The Dissertation Committee for Rajesh Paleti Ravi Venkata Durga Certifies that this is the approved version of the following dissertation:

On Integrating Models of Household Vehicle Ownership, Composition, and Evolution with Activity Based Travel Models

Committee:

Chandra R. Bhat, Supervisor

Jason Abrevaya

Ram Pendyala

Randy B. Machemehl

Stephen Boyles

On Integrating Models of Household Vehicle Ownership, Composition, and Evolution with Activity Based Travel Models

by

Rajesh Paleti Ravi Venkata Durga, B. Tech; M.S.E.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December, 2012

Dedication

To my mother Satya Seeta, and my father Ashok Babu

Acknowledgements

Before coming to the United States, I was strongly convinced that the four years as an undergraduate student in India will go down as the most memorable days in my life. Be it divine grace or sheer luck, I met great friends and equally inspiring mentors and advisors as a graduate student at UT Austin which made my stay in the US as memorable as my undergraduate life. I would like to acknowledge all the people who have helped and inspired me during these days.

First and foremost, I would like to thank my graduate supervisor Dr. Chandra R. Bhat for being such an inspirational and motivating advisor. His passion for research and teaching is unparalleled. I often carefully observed him and tried my best to imbibe these qualities. I can undoubtedly say that the past four years, working under his guidance, have been the most intellectually stimulating and academically productive days in my life. I am truly fortunate to have had the opportunity to work with him. I would also like to thank Dr. Ram Pendyala for his constant support and research collaboration. Working with two pioneers in the transportation field, both very different in their mentoring and writing styles, has helped me realize that the single most important quality that is needed to be successful is to have an inquisitive mind with passion and discipline. I am grateful to Dr. Jason Abrevaya for his unique style of teaching Econometrics which made it thoroughly enjoyable. I am also thankful to Dr. Randy Machemehl and Dr. Stephen Boyles for serving on my dissertation committee and proving valuable suggestions during the course of my research.

I would like to thank two people Dr. Naveen Eluru and Dr. Ipek Sener who were my graduate mentors at UT. Their constant support and confidence in me have been two main factors that inspired me to pursue a doctoral degree. I would like to specially thank Dr. Naveen Eluru for his advice both professionally and personally. Be it academics, playing Squash, going out to eat, watching movies, or discussing politics and philosophy, he is always there. I am grateful to Dr. Abdul Pinjari for his advice regarding career choices. Heartfelt thanks to six people in our research group at UT for their help at various stages during my research- Raghu, Marisol, Gaurav, Palvinder, Sriram, and Chrissy. I am very fortunate to have such awesome team members. Outside my research group, I would like to thank Binny and Moby for their company in the basement during late-night hours, and Dharma for his help with programming in Linux environment. I would also like to thank Lisa Macias for her help with all the administrative work and the quality control process of research publications.

I am thankful to my friends Binny, Roshan, and Raghu for Racquet Ball games; to our basement Cricket team members Binny, Moby, Gaurav, Palvinder, Brent, Chriss, Roshan, Tarun, Ravi, and Subodh; to Chrissy, Aria, and Andrea for being patient with our Cricket games; to Sriram and Saamiya for anime discussions; to Srivatsava, Vamsi, Sivakiran, Aneesh, Dharma, Sivaprasad, and Sundeep for being such great friends and making my time outside work fun and relaxing. I have learned quite a lot from all of them and will cherish their friendship for the rest of my life.

Lastly, I am very indebted to my family without whose support I would not be where I am today, and to my brother Chaitanya who is always there for me when I need him. I am eternally grateful to the God Almighty who has been the source of my strength and a constant companion in my life.

On Integrating Models of Household Vehicle Ownership, Composition, and Evolution with Activity Based Travel Models

Rajesh Paleti Ravi Venkata Durga, Ph.D.
The University of Texas at Austin, 2012

Supervisor: Chandra R. Bhat

Activity-based travel demand model systems are increasingly being deployed to microsimulate daily activity-travel patterns of individuals. However, a critical dimension that is often missed in these models is that of vehicle type choice. The current dissertation addresses this issue head-on and contributes to the field of transportation planning in three major ways. *First*, this research develops a comprehensive vehicle micro-simulation framework that incorporates state-of-the-art household vehicle type choice, usage, and evolution models. The novelty of the framework developed is that it accommodates all the dimensions characterizing vehicle fleet/usage decisions, as well as accommodates all dimensions of vehicle transactions (i.e., fleet evolution) over time. The models estimated are multiple discrete-continuous models (vehicle type being the discrete component and vehicle mileage being the continuous component) and spatial discrete choice models that explicitly accommodate for multiple vehicle ownership and spatial interactions among households. More importantly, the vehicle fleet simulator developed in this study can be easily integrated within an activity-based microsimulation framework.

Second, the vehicle fleet evolution and composition models developed in this dissertation are used to predict the vehicle fleet characteristics, annual mileage, and the

associated fuel consumption and green-house gas (GHG) emissions for future years as a function of the built environment, demographics, fuel and related technology, and policy scenarios. This exercise contributes in substantial ways to the identification of promising strategies to increase the penetration of alternative-fuel vehicles and fuel-efficient vehicles, reduce energy consumption, and reduce greenhouse gas emissions.

Lastly, this research captures several complex interactions between vehicle ownership, location, and activity-travel decisions of individuals by estimating 1) a joint tour-based model of tour complexity, passenger accompaniment, vehicle type choice, and tour length, and 2) an integrated model of residential location, work location, vehicle ownership, and commute tour characteristics. The methodology used for estimating these models allows the specification and estimation of multi-dimensional choice model systems covering a wide spectrum of dependent variable types (including multinomial, ordinal, count, and continuous) and may be viewed as a major advance with the potential to lead to redefine the way activity-based travel model systems are structured and implemented.

Table of Contents

List of Tables

List of Figures

CHAPTER 1: Motivation

1.1 Introduction

In the U.S., energy-related activities account for three-quarters of total human-generated greenhouse gas (GHG) emissions, mostly in the form of carbon dioxide (CO_2) emissions from burning fossil fuels. Recent projections show that the nation's CO_2 emissions will increase from about 5.9 million metric tons in 2006 to 7.4 million metric tons in 2030 if measures are not taken to reduce carbon emissions (NAS, 2008). The transportation sector accounts for 27 percent of all greenhouse gas emissions, ranking second only to electricity/power generation (EPA, 2010; see Figure 1.1). The contribution of the transportation sector to greenhouse gas emissions stems largely from the burning of fossil fuels that can be associated with vehicular travel (EPA, 2010). While about one-half of the nation's CO_2 emissions originate from large stationary sources such as power plants, the transportation sector is one of the most rapidly rising sources of GHG emissions. For example, total U.S. GHG emissions rose 13% between 1990 and 2003. In comparison the contribution of the emissions from the transportation sector rose 24% during the same period (EPA, 2006).

Transportation continues to be a dominant component of the world's energy consumption, accounting for 30 percent of the world's energy use and 95 percent of global oil consumption (EPA, 2010). In the United States, transportation accounts for about 19 million barrels or 60 percent of all petroleum consumption per day, a statistic that has increased by more than 50 percent over the past 35 years.

1.2 Importance of Vehicle Type Choice and Usage Modeling

Within the transportation sector, on-road vehicular travel accounts for a substantial portion of GHG emissions. In the category of on-road vehicular travel-based GHG emissions, passenger cars and light duty trucks (SUVs, pickup trucks, vans and minivans) are the largest sources, accounting for nearly two-thirds of the emissions attributable to vehicular travel (EPA, 2006).

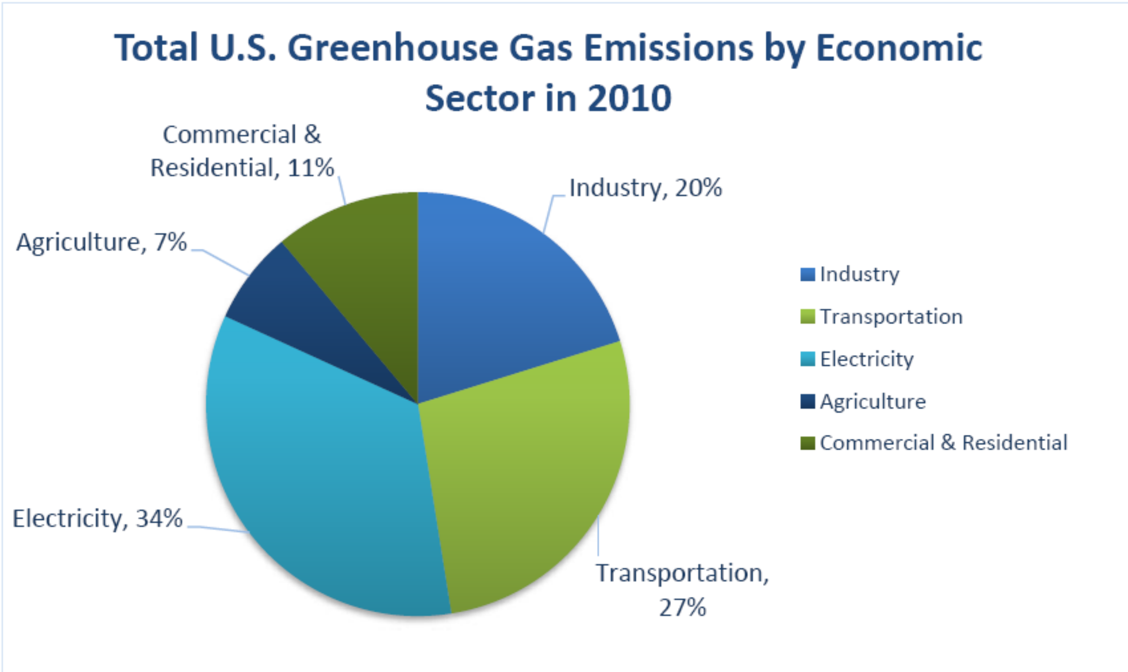

Figure 1.1 Greenhouse Gas Emissions by Sector[1]

Between 1990 and 2003, while emissions from passenger cars increased by just about 2%, GHG emissions from light duty trucks (LDTs) increased by about 50%. The increase in GHG emissions from automobiles and LDTs reflects the overall growth in travel demand (measured in vehicle miles of travel or VMT) and the substantial shift in household vehicle fleet composition towards larger, less fuel-efficient vehicles (see Bhat and Sen, 2006). The SUV market share, in particular, increased from just about 1% in 1976 to over 25% in 2003, while passenger cars experienced a decrease in share from over 80% to just about 47% during this period (EPA, 2006). This trend has been reversing somewhat in the past few years (see Bhat *et al.*, 2009). However, the fact remains that there is a substantial vehicle mix by body type and make/model on roadways. A related dimension is the fuel type of the vehicles on the road. Nearly all vehicles in use today are powered by fossil fuel derived from crude oil. However, over the next 15 to 20 years, and driven by fossil fuel independence and global climate change

[1] Source: Modified from Sources of Greenhouse Gas Emissions (EPA 2011)

considerations, energy fuel composition and power sources are likely to change and the number of vehicle fuel options for consumers is expected to increase significantly. Conventional gasoline will likely be blended with plant-derived ethanol or possibly substituted with high-ethanol-content blends such as E85 (85% ethanol); petroleum diesel may also be blended with or completely replaced by biodiesel, which can be derived from a variety of biological sources; and with improvements in technology rapidly being introduced, electric powered vehicles are expected to play a larger role. Another important dimension relevant to GHG emissions within any given vehicle body type/make/model and fuel type is the age (vintage) of the vehicle, with older vehicles contributing more to GHG emissions relative to newer, more fuel-efficient, vehicles. The impact of the composition and utilization of the household vehicular fleet on energy consumption and greenhouse gas emissions calls for the incorporation of behavioral models of vehicle type choice and utilization in transportation demand forecasting models. Such models would provide the ability to forecast energy and environmental impacts of shifting vehicle ownership and utilization patterns arising from alternative policy decisions, the advent of new alternative fuel vehicle technologies, and changes in household and personal vehicular preferences.

1.3 Integration with Activity-Based Travel Demand Models

In the context of improved tools and methods, the movement from the trip-based approach to the activity-based approach has resulted in new microsimulation models of activity-travel demand that simulate individual traveler activity-travel patterns along the continuous time axis. However, there has been little parallel work to develop microsimulation models of vehicle fleet evolution and composition, and embed these within larger forecasting systems of travel and emissions. This lack of progress may be traced to three impediments that have hampered vehicle fleet prediction efforts. The first impediment has been the very limited data available to analyze vehicle acquisition choices and the potential penetration of new vehicle and fuel types. Cross-sectional data of what households choose today may be used to develop projection models into the future, but such models will inevitably not reflect choices in a rapidly changing vehicle type market. The same is true even if household panel data were available to show dynamic vehicle transactions over the past several years.

The second issue is that the methodology used to analyze vehicle type and usage has been very limiting in terms of modeling the gamut of household vehicle type choices of the entire fleet owned by the household at any given time as well as the evolution of the household vehicle fleet over time. Earlier studies focus on a single vehicle (such as the vehicle type of the most recently purchased vehicle, or on the most used vehicle), with the characterization of vehicle type based on a narrow definition of aggregate body type (such as a car or an SUV). Such modeling approaches are inadequate for analyzing the spectrum of household vehicle choices and fleet evolution decisions (including multiple vehicles, body type/make/model, fuel type, which vehicles to replace and when, how to dispose vehicles, and if and when to add vehicles). As a result, they are extremely limited in their ability to inform the design of proactive land-use, economic, and transportation policies to influence vehicle fleet composition in a way that can reduce crude oil dependence, air quality problems, and GHG emissions.

The third impediment has been the perceived difficulties in embedding a vehicle fleet composition model within an activity-based modeling framework. In most of the existing activity-based travel demand models, the impact of vehicle ownership on daily activity-travel patterns of an individual is modeled by using the overall count of vehicles owned by the household in which the individual resides as an explanatory variable in the models used to predict the daily-activity travel choices of the individual. The important dimensions of the body type, fuel type, and vintage of the vehicles owned by the household are completely ignored. Consequently, in current practice, where travel demand models are interfaced with EPA's MOBILE6 or the recently released MOVES model or the EMFAC model in California for emissions forecasting, default values are often used to represent the mix of vehicles that contribute to vehicle miles of travel (VMT), the age distribution of the vehicle fleet, and the mix of VMT by type of roadway. Although the use of these default values offers simplicity, they may not reflect local conditions with respect to vehicle fleet composition, and even if they do, there is no basis to forecast future vehicle fleet composition in response to changes in such factors as fuel prices, socio-economic shifts (*e.g.*, aging of the US population), and policy decisions (*e.g.*, allowing vehicles attaining a certain fuel efficiency to use high-occupancy vehicle or HOV lanes).[2] In fact, recent automotive sales figures clearly show the shift towards smaller and more fuel efficient vehicles in response to the recent run-up in gasoline prices and growing public sensitivity to the environment. Even small changes in vehicle fleet mix have the potential to result in large changes in emission totals and it is therefore of utmost importance and interest to forecast vehicular fleet composition accurately using models that offer sensitivity to a host of socio-economic, policy, and modal conditions and integrate these models with activity-based travel demand models.

1.4 Integrated Modeling of Vehicle Ownership, Location, and Activity Travel Choices

Activity-based model systems are increasingly being deployed to microsimulate daily activity-travel patterns of individuals. There are a host of tour attributes of interest that are modeled within these systems. However, a dimension that is often missed is that of vehicle type choice, a variable of considerable importance in the energy consumption and emissions estimation arena. Another issue that arises is that most tour attributes are modeled independently or sequentially with loose coupling across the models, thus ignoring important endogeneity effects that may exist across multiple tour dimensions.

Of equal significance, the multitude of location and activity-travel choices made by people are interconnected in complex ways (Lerman, 1976). There are long term location choices such as residential location, work location, and school location choice. There are medium term choices such as vehicle ownership, and there are shorter term travel choices such as activity participation, time of day choice, trip chaining, trip sequencing, destination choice, mode choice, and route choice. While it may be convenient from an

[2] Although the FHWA offers some guidance on how default values on vehicle mix distributions can be adjusted using local vehicle registration data and vehicle classification counts (http://www.fhwa.dot.gov/environment/conformity/emission/emismeth7.htm), these values are still aggregate-level numbers that offer little for forecasting future vehicle fleet composition.

operational model development perspective to assume that these choices are made in a neat sequential fashion – with long term choices affecting medium term choices, both long and medium term choices affecting short term choices, and short term choices linked together in a series – there is often little behavioral basis underlying such sequential decision frameworks.

On the contrary, there is a growing and important body of evidence that supports the notion that people make a multitude of choices as a "bundle", choosing a series of location and activity-travel attributes that define their lifestyle jointly. Thus, there is a need to simultaneously model a multitude of choices in an integrated framework. However, the specification and estimation of such simultaneous equations model systems has remained a challenge and prevented progress in this domain.

1.5 Current Research

Within the context of improved household vehicle fleet modeling and integration with activity-based travel models, the current dissertation aims to contribute to the transportation field in five major ways.

The first objective is to develop a comprehensive vehicle micro-simulation framework that can be gainfully and efficiently incorporated within an activity-based microsimulation framework. As a part of this effort, state-of-the-art household vehicle type choice, usage, and evolution models accommodating all the dimensions characterizing vehicle fleet/usage decisions, as well as vehicle transaction decisions (i.e., fleet evolution) over time will be estimated.

The second objective of this research is to develop a microsimulation platform to apply the modeling system developed in this research to predict vehicle fleet and usage and the associated fuel consumption and GHG emissions for future years as a function of the demographics, fuel and related technology, and policy scenarios.

The third objective is to further enhance the vehicle type choice models to accommodate spatial interactions among households in both the choice of the number of vehicles owned as well as the type of vehicles owned. The spatial vehicle type choice model will be used to demonstrate that ignoring the spatial dependency effect results in elasticity estimates that are substantially different and lower in magnitude –presumably because changes in behavior that are brought about by household interactions in space are not taken into account. The findings have important implications for model development and application in the policy forecasting arena.

The fourth objective of this study is to capture the interactions between vehicle ownership and activity-travel decisions of individuals by developing an integrated modeling framework of several daily activity-travel choices and vehicle type choice. The methodologies developed in this study that allow the specification and estimation of complex multi-dimensional choice model systems in simultaneous equations frameworks may be viewed as a major advance with the potential to lead to significant breakthroughs in the way activity-based travel model systems are structured and implemented.

The final objective of this research is to extend the integrated modeling framework to develop an econometric model system that simultaneously considers six different choice dimensions that cover disparate temporal scales from the long term location choices to the short term activity-travel choices (with vehicle ownership being a medium-term choice), and include a variety of dependent variable types in a unifying framework. The six dimensions include residential location choice, work location choice, auto ownership, commuting distance, commute mode, and number of stops on commute tours. The model specification accounts for the possible presence of correlated unobserved factors affecting multiple choice dimensions and thus reflects the influence of self-selection effects that can have important implications for policy forecasts.

1.6 Dissertation Outline

The rest of the dissertation proposal is structured as follows. Chapter 2 presents earlier literature related to vehicle type choice and usage analysis, spatial vehicle type choice modeling, and integrated modeling of vehicle type/ownership, location, and daily activity-travel choices. Furthermore, Chapter 2 also positions the current research in the context of these earlier studies reviewed. Chapter 3 describes the vehicle fleet microsimulation framework developed in this study, the methodological details of the models used to analyze the vehicle composition and evolution decisions, an overview of the data set used for estimating the models in the microsimulation framework and the model estimation results. Chapter 4 describes the microsimulation platform, the forecasting procedure, and the prediction results under different policy scenarios. Chapter 5 presents the methodological details of the spatial vehicle type choice model and the empirical results of such a model developed for the California region. Chapter 6 presents the multi-dimensional modeling methodology used to jointly model daily activity-travel choices and vehicle type choice decisions, the data used for estimating the integrated model and the results of the empirical analysis. Chapter 7 discusses the unifying econometric model system that can analyze a variety of dependent variable types and the empirical results of an integrated model of residential and work location decisions, vehicle ownership, and several commute choices. Finally, Chapter 8 summarizes the work accomplished and outlines a plan for future research.

CHAPTER 2: Earlier Research and the Current Study in Context

This chapter provides a detailed review of earlier work undertaken relevant to the two main objectives of the dissertation identified in the previous chapter. The literature reviewed is grouped under two headings- 1) Earlier research in vehicle fleet composition and evolution, and 2) Earlier research in integrated modeling of activity-travel choices. Both the substantive and methodological contributions of these earlier studies have been documented. Specifically, Section 2.1 presents an overview of the earlier studies which modeled vehicle type and usage as well as vehicle transaction decisions and Section 2.2 discusses earlier literature on spatial vehicle type choice modeling. Section 2.3 provides an overview of the existing literature focusing on development of integrated models of multiple activity-travel choices and vehicle type choice while Section 2.4 extends the discussion to earlier literature on integrated location, vehicle ownership, and activity-travel choices. Finally, Section 2.5 outlines the contributions of the current dissertation in the context of the studies reviewed.

2.1 Vehicle Composition and Evolution

In light of global energy consumption and emissions concerns, several studies in the recent past have focused attention on the types of vehicles owned by households – the type of vehicle being defined by some combination of body type or size, fuel type, and the age of the vehicle – as well as the mileage (utilization) of the vehicles (for example, see Bhat *et al.*, 2009, Brownstone and Golob, 2009). These studies explicitly recognize that energy consumption and GHG emissions are not only dependent on the number of vehicles owned by households, but also on the mix of vehicle types and the extent to which different vehicle types are utilized (driven).

The literature has recognized for a long time, however, that household vehicle ownership (or fleet composition and utilization) models are only capable of providing a snapshot of vehicle holdings and mileage, as such models are routinely estimated on cross-sectional data sets that offer little to no information on vehicle transactions over time (Hensher and Plastrier, 1985, Jong and Kitamura, 1992). As the focus of transportation planning is largely on forecasting demand over time, it is desirable to have a vehicle fleet evolution model that is capable of evolving a household's vehicle fleet over time (say, on an annual basis) by analyzing the dynamics of vehicle transaction decisions over time. In addition, the vehicle evolution model system should be sensitive to a range of socio-economic and policy variables to reflect that vehicle transaction decisions are likely influenced by the types of vehicle technologies that are and might become available, public policies and incentives associated with acquiring fuel-efficient or low/zero-emission vehicles, and household socio-economic and location characteristics (Brownstone *et al.*, 2000, Haan *et al.*, 2009, Mueller and Haan, 2009).

Unfortunately, however, the development of dynamic transactions models has been hampered by the paucity of longitudinal data on vehicle transactions that inevitably occur over time. Mohammadian and Miller (2003) use about 10 years of data to model vehicle ownership by type and transaction decisions over time, but do not include fuel type as one of the attributes of vehicles. Yamamoto *et al.* (2008) use panel survey data to model vehicle transactions using hazard-based duration formulations as a function of changes in household and personal demographic attributes. Their study also shows the role of history dependency in vehicle transaction decisions with a preceding decision in time affecting a subsequent transaction decision. Two other studies in the recent past-Prillwitz *et al.* (2006) and Yamamoto (2008) focused on the impact of life course events on car ownership patterns of households using panel data. Prillwitz *et al* (2006) estimated a binary probit model to analyze the increase in car ownership level (1 corresponding to an increase and 0 otherwise) using German Socioeconomic panel data from 1998 to 2003, while Yamamoto (2008) developed hazard-based duration models and multinomial logit models to analyze the vehicle transaction decisions using panel data in France and retrospective survey data for Japan respectively. These studies of dynamic vehicle transactions behavior emphasize the need for simulating vehicle fleet composition and utilization over time to accurately estimate energy consumption and GHG emissions arising from human activity-travel choices. However, because of the difficulty of collecting data over time (including costly design/implementation of panel surveys and survey attrition over time; see Bunch, 2000), dynamic models have focused primarily on vehicle ownership (*i.e.*, transactions) with inadequate emphasis on the vehicle type, usage, and vintage considerations of the household fleet. Further, in today's rapidly changing vehicle market, a substantial limitation of panel models based solely on revealed choice data is that these models do not consider the range of vehicle, infrastructure, and alternative fuel advances on the horizon, and are thus insensitive to technological evolution.

2.2 Spatial Modeling of Vehicle Type Choices

The past decade has seen increasing attention being paid to accommodating spatial dependency effects in modeling choice-making behaviors of agents in a variety of contexts (Anselin, 2010). There have been several efforts in the recent past to apply spatial correlation structures that have been developed for modeling continuous dependent variables in the context of discrete choice models of behavior (see recent reviews of this literature in Franzese *et al.*, 2010; Anselin, 2010; Bhat *et al.*, 2010). However, these efforts have been hampered by the need to evaluate multidimensional integrals of the order of the product of the number of decision agents and the number of alternatives minus one for unordered multinomial response choice models. Traditional simulation-based estimation methods as well as the Bayesian Markov Chain Monte Carlo (MCMC)-based estimator (LeSage and Pace, 2009) are very cumbersome, if not infeasible, to implement in typical empirical contexts with even moderate sample sizes and choice sets (Bhat, 2011; Franzese *et al.*, 2010). The result is that researchers have settled on imposing restrictive localized spatial dependence that provides computational tractability at the expense of behavioral representation. For example, it has been assumed that agents within a certain spatial region have a constant spatial dependence that "drops off a cliff" at the boundary of the region, with no spatial dependence between neighboring decision agents that fall on either side of the boundary (Dugundji and Walker, 2005). Others have simplified the problem by employing restrictive aggregate-level spatial clusters, grouping decision agents into much fewer regional clusters, and imposing a proximity-based spatial dependence pattern at the regional cluster level (Phaneuf and Palmquist, 2003; Smith and LeSage, 2004). While such approaches lead to tractability in estimation, the basic problem with these restrictive specifications is that space is considered discrete, while space is, in reality, a continuous entity. As a consequence, these earlier studies are more susceptible to the modifiable areal unit problem (MAUP) than studies that accommodate global spatial dependency effects at the unit of the decision agents (Páez and Scott, 2004).

Several other studies (*e.g.*, Mohammadian *et al.*, 2005; Adjemian *et al.*, 2010) have similarly side-stepped the intractability problem inherent in global and general spatial dependency structures by assuming that the dependency originates only from observed exogenous covariates of proximate decision agents. However, this is rather untenable in the context of several choice situations where the spatial dependence naturally arises from didactic interactions between decision agents. To elucidate, households may be viewed as developing utilities (or preferences) for vehicle type choice alternatives based on a set of observed factors (such as income and presence of children in neighboring households) as well as unobserved tastes, attitudes, and location factors (such as how "green" a household is in its views and whether there are continuous sidewalks/bicycle paths in the neighborhood). The utility vector of one household is likely to be influenced by the utility vector of other nearby households due to didactic interactions and interchanges (where utility signals get bounced around across decision agents). In this process, there is a "spatial spillage" effect not only based on the observed covariate effects of neighboring households, but also due to unobserved factors. For example, a neighboring household's perception of "greenness" or the quality of sidewalks/bike paths may spill over and influence choices of another household. Further, there may be residential self-selection effects leading to a sorting of households based on similarity in unobserved vehicle type choice preferences.

In discrete choice models, ignoring these spillage effects due to observed factors and/or due to unobserved factors will, in general, lead to inconsistent estimates of the effects of observed covariates. As indicated by Anselin (2003), it behooves the analyst to include spatial “spillover” effects in both the observed covariates as well as the errors unless there are strong a priori reasons not to do so.

2.3 Integrated Modeling of Vehicle Type & Activity-Travel Choices

There are a variety of tour attributes of interest in the context of designing and implementing activity- or tour-based microsimulation models of travel. Tour-based model systems generally involve the modeling of all or a subset of daily activity-travel choices including tour type, number of intermediate stops, time of day choice, mode choice, intermediate stop purpose, number of individuals on the tour, destination choice for primary and secondary stops, and activity episode duration (Bhat *et al.*, 2004; Vovsha and Bradley, 2006; Bowman and Ben-Akiva, 2001). Model components pertaining to the various choices and dimensions of interest are often linked together to form a sequential chain of models, with potential feedback involving logsums for choice variables where nested logit model forms are used (Wen and Koppelman, 1999). While the above structures are certainly convenient from a model deployment and application standpoint, they are limited in their ability to simultaneously model the complex inter-relationships among the multiple tour attributes while accounting for the possible presence of correlated unobserved attributes across choice dimensions.

The development of simultaneous equations models of activity-travel behavior has been of much interest in the travel behavior research domain for decades for precisely this reason (e.g., Mannering and Hensher, 1987; Kitamura *et al.*, 1996; Pendyala and Bhat, 2004). Simultaneous equations modeling has been motivated by the desire to appropriately represent endogeneity in choice processes where correlated error structures may exist, and thus make travel behavior models more accurately capture behavioral processes at play. Ignoring endogeneity that may exist across choice dimensions that are inter-related with one another results in coefficient estimates that are inconsistent and biased (Mokhtarian and Cao, 2008), with inevitable adverse impacts on the quality of the forecasts provided by such models.

From a methodological perspective, the profession has been limited by the complexity associated with formulating and estimating simultaneous equations models that capture a multitude of choice dimensions in a joint model system. Most simultaneous equations models have been limited to bivariate model systems (e.g., Hamed and Mannering, 1993; Bhat, 1998; Yamamoto and Kitamura, 1999; Golob, 2000; Ye and Pendyala, 2007), either involving two discrete choice variables or a combination of discrete and continuous choice variables. While these models have undoubtedly provided key insights into endogeneity of choice processes, the inability to model more than two choice dimensions simultaneously has made it difficult to account for endogeneity across a multitude of choices that may be made as a package or bundle (Chung and Rao, 2003). The complexity associated with estimating larger multidimensional choice models systems with a mixture of endogenous variable types using classical econometric formulations has led to a stream of literature utilizing structural equations methods (Golob, 2003; Bagley and Mokhtarian, 2002). In structural equations models, multiple endogenous variables may be modeled simultaneously while accounting for the possible presence of significant error covariances. These models have been able to shed considerable light on the complex interactions across multiple activity-travel variables; however, the key issue associated with structural equations models is that they cannot accommodate multinomial choice variables – which happen to be one of the most important variable types in travel modeling (for example, destination choice, mode choice, time of day choice, and activity type choice).

More recently, progress has been made in the multidimensional modeling of choice processes in the activity-travel arena (Pinjari *et al.*, 2011; Eluru *et al.*, 2010). These efforts exploit some of the dramatic advances in choice model specifications and estimation methods that have occurred in the recent past (Bhat and Eluru, 2010; Bhat *et al.*, 2008). These advances make it possible to formulate model specifications that account for complex observed and unobserved inter-relationships that exist among multiple dependent variables and to estimate such model systems without having to resort to simulation-based approaches that quickly become computationally burdensome and potentially imprecise as the dimensionality of the problem increases (Bhat, 2011).

2.4 Integrated Modeling of Location, Vehicle Ownership, & Activity-Travel Choices

The evidence in favor of attempting to model a multitude of choice dimensions simultaneously in a joint modeling framework is quite irrefutable and growing (Abraham and Hunt, 1997). Notably, the body of work examining the impact of land use measures on travel behavior suggests that there are considerable self-selection effects wherein households tend to locate in neighborhoods that have attributes consistent with their lifestyle and mobility preferences (Bhat and Guo, 2007; Cao *et al.*, 2008). For example, households that are not auto-oriented choose to locate in transit and pedestrian friendly neighborhoods that are characterized by mixed and high-density land use. If that is the case, then it is likely that the choices of residential location, vehicle ownership, and commute mode choice (for example) are being made jointly as a bundle rather than in a sequential fashion in which residential location choice is chosen first, residential location affects vehicle ownership (which is chosen second), and vehicle ownership affects commute mode choice (which is chosen third). The sequential model is likely to over-estimate the impacts of residential location (land use) attributes on activity-travel behavior because it ignores self-selection effects wherein people who locate themselves in such neighborhoods were auto-disoriented to begin with. These lifestyle preferences and attitudes constitute unobserved factors that simultaneously impact long term location choices, medium term vehicle ownership choices, and short term activity-travel choices; the only way to accurately reflect their impacts and capture the "bundling" of choices is to model the choice dimensions together in a joint simultaneous equations modeling framework that accounts for correlated unobserved lifestyle (and other) effects.

There is a large body of work on simultaneous equations modeling in location and activity-travel choices with a view to better understand the bundling of choice behaviors while addressing the challenges associated with estimating such econometric model systems. The formulation, specification, and estimation of multi-dimensional choice model systems in which there are a variety of dependent variable types (continuous, ordinal, multinomial, count) has proven to be a challenging task because of the need to evaluate large multi-dimensional integrals of mixtures of distributions in such model systems. As a result, a number of papers in this domain have limited the number of choice dimensions considered to two or have adopted alternative approaches (such as structural equations modeling methods which cannot adequately handle multinomial choice variables) to estimate models with more than two dependent variables. The recognition of simultaneity in choice making behaviors has its roots in microeconomic consumer choice theory as evidenced by the partial or general equilibrium class of models developed by LeRoy and Sonstelie (1983) who investigated relationships between residential choice, income, and mode choice, Brown (1986) who postulated that residential location and commute travel mode are goods that consumed simultaneously, and DeSalvo and Huq (1996, 2005) who jointly model residential location, income, and commute mode choice.

In the transportation domain, examples of simultaneous equations models of location and activity-travel choice behaviors abound. Bagley and Mokhtarian (2002) specify and estimate a nine-equation structural equations model system to explore relationships across residential location, travel choices, work location, and attitudinal variables. Choo and Mokhtarian (2004) also explore the influence of attitudinal variables on traveler choices by focusing on vehicle type choice. Attitudinal variables, that are often unobserved, play an important role in shaping a multitude of choices, thus calling for the bundling of choices in a simultaneous equations framework where such correlated unobserved factors can be adequately reflected. Van Acker and Witlox (2010a, 2010b) also use structural equations modeling approaches to explore relationships between built environment attributes and vehicle use in a simultaneous equations modeling framework. Vance and Hedel (2007) model the choice of driver status and vehicle use (distance traveled) simultaneously using an instrumental variables approach. Vega and Reynolds-Feighan (2009) employ a cross-nested logit model to study the simultaneous choices of residential location and travel mode under two scenarios of employment (central city versus suburb). Ye *et al.* (2007) use a bivariate probit modeling framework to examine the relationship between trip chaining and mode choice, while Konduri *et al.* (2011) employed a probit-based joint discrete-continuous model to tie vehicle type choice and tour length (distance) together. Brownstone and Golob (2009) used Bayesian estimation approaches to jointly analyze residential location choice in the context of vehicle type choice and usage and find significant presence of endogeneity in the choice dimensions examined. A similar study was undertaken by Eluru *et al.* (2009), except that they employed Copula-based estimation approaches. Krizek (2003) introduces a tour-based framework to analyze relationships jointly among neighborhood access, number of tours, tour type, and tour distance.

More recently, Eluru *et al.* (2010) and Pinjari *et al.* (2011) constitute key efforts to build integrated multi-dimensional choice models that tie longer term location choices and shorter term activity-travel choices together. The model system in the former study is estimated using Copula-based frameworks while the model system in the latter study is estimated using more traditional simulation-based estimation approaches. Both of these studies showed strong evidence of the bundling of choices with correlated unobserved effects. Many of the studies cited in this section have noted the computational challenges associated with estimating multi-dimensional choice models, particularly in the presence of a mixture of dependent variable types. However, recent advances in estimation methods, and in particular, the emergence of the Maximum Approximate Composite Marginal Likelihood (MACML) approach (Bhat, 2011), have provided the much needed computational breakthroughs required to estimate multi-dimensional choice model systems and bring them closer to modeling practice.

In summary, there is ample evidence of the interest in, importance of, and need for modeling a multitude of location and activity-travel choice dimensions across temporal scales. This study is aimed at overcoming a number of limitations associated with previous work in this arena. First, several studies examine only a limited number of choices (usually, just two dimensions) within a single temporal frame (*e.g.*, Waddell *et al.*, 2007). Second, several studies attempt to build relationships across temporal scales, but do so again for very limited number of dimensions (*e.g.*, Brownstone and Golob, 2009; Bhat and Guo, 2007). Third, several studies employ the structural equations modeling approach to tie a multitude of dimensions across time scales together, but these approaches cannot accommodate multinomial choice dependent variables (such as mode choice). Fourth, studies such as those by Eluru *et al.* (2010) and Pinjari *et al.* (2011) employ estimation approaches that either are burdensome or involve numerical approximations that may compromise the properties of parameter estimates.

2.5 Current Study in Context

2.5.1 Comprehensive Vehicle Fleet Composition and Evolution Framework

This study develops a comprehensive vehicle fleet composition, utilization, and evolution framework that can be integrated within activity-based microsimulation models of travel demand. The model includes several components that allow one to not only predict current (baseline) vehicle holdings and utilization (by body type, fuel type, and vintage) but also simulate vehicle transactions (including addition, replacement, or disposal) over time. The usual data limitation is overcome in this study through the use of a unique large sample survey data set collected recently in California. Specifically, the survey not only included a revealed choice (RC) component of current vehicle holdings and vehicle purchase history, but also a stated intentions (SI) component related to intended vehicle transactions in the future and a stated preference (SP) component eliciting information on vehicle type choice preferences. By pooling data from these components, a range of vehicle types (including those not commonly found in the market place) could be included in a vehicle type choice model. The resulting model system can be used to test the effects of a range of policy variables on vehicle fleet composition, utilization, and evolution decisions.

2.5.2 Spatial Vehicle Type Choice

This dissertation aims to contribute to the vehicle ownership and fleet composition analysis literature by presenting a multinomial probit model that explicitly accounts for spatial interaction effects in these choice phenomena. Underlying the multinomial probit model with spatial interaction effects is a behavioral framework that not only estimates the number of vehicles owned by a household, but also the vehicle type choice – thus allowing the construction of the entire vehicle fleet for a household, while explicitly considering spatial dependency effects. In the current study, a spatial lag formulation is adopted to accommodate global spatial dependence effects (due to both observed covariate and error spillage effects) in household vehicle type choice decisions. The specific model structure and formulation implemented in this study allows the modeling of the entire vehicle fleet composition of households. The development of a multinomial probit model with continuous spatial dependency effects (due to both observed and unobserved factors) constitutes the novel contribution of this dissertation.

2.5.3 Integrated Tour Based Model System

Of equal significance, the dissertation aims to further advance the development and estimation of multidimensional choice model systems of activity-travel behavior by considering a bundle of endogeneous variables that characterize tours in activity-based travel model systems. The four attributes considered in this study are tour complexity, passenger accompaniment, vehicle type choice, and total tour length. While this set of dimensions is certainly not exhaustive by any means, it does represent an important group of choices from a transportation modeling and planning perspective that are likely to be inter-related to one another.

Within the context of the emerging energy sustainability and greenhouse gas emissions reduction debates, the modeling of vehicle type choice and tour length is of particular interest as these choices directly impact energy and environmental outcomes. Despite the importance of vehicle type choice in this arena, rarely has vehicle type choice been explicitly modeled in tour-based models. Modeling and tracking vehicle type choice within the larger context of activity-based models can greatly inform emissions inventory models that are able to take advantage of detailed information of vehicle trajectories by type of vehicle. Moreover, from a policy perspective, one can examine the potential (sometimes, unintended) consequences of actions. For example, consider a scenario where rebates are instituted for the purchase of fuel efficient vehicles to enhance their presence in the fleet. Individuals can, however, travel farther distances using more fuel efficient vehicles at the same cost as they would travel shorter distances with gas guzzling vehicles. Then, the longer travel distances induced by the acquisition of fuel efficient vehicles (spurred by the policy actions) would, at least in part, negate the benefits associated with encouraging fuel efficient vehicle acquisition in the population. In addition, total vehicle miles of travel could increase, leading to greater levels of congestion and delay. It is these types of complex inter-relationships that can be captured through the estimation and deployment of multi-dimensional choice model systems.

2.5.4 Integrated Residential Location, Work Location, Vehicle Ownership, & Commute Tour Characteristics

This dissertation attempts to overcome the limitations associated with previous work in the specification and estimation of multi-dimensional model systems of location and activity-travel choices. In this study, six choice dimensions are tied together in a joint modeling framework. Residential location and workplace location choices are long term multinomial choice variables, commute distance (which is an outcome of residential location and workplace location choices) is a long term continuous variable, household vehicle ownership is a medium term ordinal dependent variable, commute mode choice is a short-term multinomial travel choice variable, and finally, number of stops made during commute tour is an ordinal dependent variable. These six variables are tied together in a temporal framework while recognizing the bundling of these choice dimensions associated with the jointness or simultaneity in decision-making. The study aims to contribute substantially to the literature by presenting empirical insights into the relationships across choice dimensions and offering a methodological approach that makes the estimation of such model systems feasible.

CHAPTER 3: Vehicle Fleet Composition and Evolution Framework

The material in this chapter is drawn substantially from the following published paper:

Paleti, R., N. Eluru, C.R. Bhat, R.M. Pendyala, T.J. Adler, and K.G. Goulias (2011) Design of comprehensive microsimulator of household vehicle fleet composition, utilization, and evolution. *Transportation Research Record*, 2254, 44-57.

This chapter discusses the vehicle fleet microsimulation framework as well as provides the methodological details of all the models that constitute this framework. Specifically, Section 3.1 describes two main components of the vehicle framework- the vehicle selection module and the vehicle evolution module. Specific details regarding the merits of the microsimulation framework and how it addresses the limitations of earlier studies reviewed in Chapter 2 are also be provided. Section 3.2 discusses the copula-based joint modeling and the dynamic vehicle transaction modeling structures used to analyze the vehicle type/usage and vehicle transaction decisions in the microsimulation framework developed. Section 3.3 describes the three components of the California Vehicle Survey data, the data processing, and sampling procedures. Section 3.4 discusses the empirical results of the models estimated in the vehicle selection and evolution modules, model fit measures, and results of validation exercise undertaken on a small sample excluded during estimation. Section 3.5 concludes the chapter by summarizing the major contributions and findings of the study.

3.1 Vehicle Fleet Composition and Evolution Framework

Figure 3.1 presents the vehicle fleet composition and evolution framework used in the current study. First, there is a base year (baseline) model capable of predicting the current vehicle fleet composition and utilization of a household. In order to recognize the fact that the vehicles owned by a household at any given point in time are not acquired contemporaneously, the household is deemed to have acquired the vehicles on multiple choice occasions. Based on extensive analysis of travel survey data sets, it has been found that the number of vehicles owned by a household is virtually never greater than the number of adults in the household plus two (in the data set used in the current analysis, 99.7% of households were covered by the condition that the number of vehicles is no greater than the number of adults plus two; note also that the approach is perfectly generalizable to the case where the number of vehicles is never greater than the number of adults plus K, where K is any positive integer determined by the analyst based on the data being studied). Then, each household is assumed to have a number of "synthetic" choice occasions (on which to acquire a vehicle) equal to the number of household adults plus two. In the figure, an example is shown for a two-adult household with four possible choice occasions. In each choice occasion, a household may acquire a vehicle and associate an amount of mileage (utilization) to it, or may not acquire a vehicle at all. Further, since the temporal sequence of the purchase of the vehicles owned by the household is known, the impacts of the types of vehicles already owned on the type of vehicle that may be purchased in a subsequent purchase decision could be accommodated. This mimics the dynamics of fleet ownership decisions. Once the base year fleet composition and utilization has been established for each household, the simulator turns to the evolution component. The evolution component works on an annual basis with households essentially faced with a number of possible choice alternatives (decisions). For each vehicle in the household, a household may choose to either dispose the vehicle (without replacing it) or replace the vehicle (involving both a disposal and an acquisition). If the choice is to replace the vehicle, then the vehicle selection module model estimation results can be applied to determine the type of vehicle that is acquired and the mileage that is allocated to it. Finally, a household may also choose to add a net new vehicle to the household fleet. In the case of an addition, once again the vehicle type choice and utilization model from the first simulator component can be applied to the vehicle acquired. Note that this framework overcomes the limitations of past studies that generally allowed only one possible transaction in any given year. Further, dependency between transaction decisions can be accommodated by including the number of years since an earlier transaction decision. For example, a vehicle may be less likely to be replaced if another vehicle was replaced the year before or if a vehicle was added the year before. Similarly, a vehicle may be less likely to be added if a vehicle was added the year before or if another vehicle was replaced the year before.

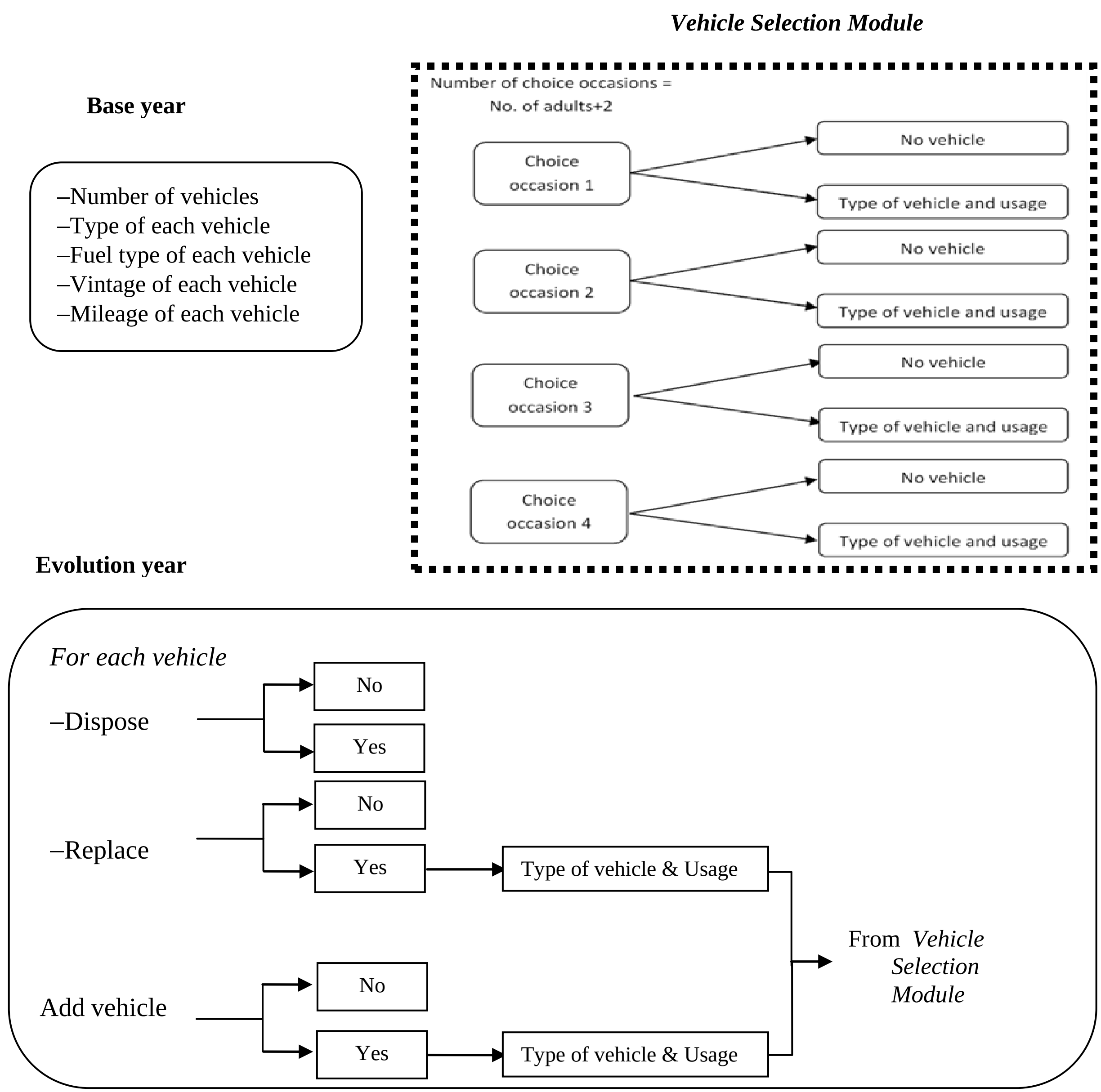

Figure 3.1Vehicle Fleet Composition, Utilization, and Evolution Simulator Framework

3.2 Methodology

3.2.1 Vehicle Selection Module

The vehicle selection module employs the traditional discrete-continuous framework for modeling the base year vehicle fleet composition and utilization. The vehicle fleet is described by a multinomial logit model of vehicle body type, fuel type, and vintage, and mileage (in logarithmic form) is modeled using a linear regression model. The methodology is the same as that described in Eluru *et al.* (2010). As discussed earlier in Section 3.1, the vehicle fleet and usage decisions are assumed to occur through a series of unobserved (to the analyst) vehicle choice occasions, with the number of vehicle choice occasions being equal to N+2 (N being the number of adults in the household).

Let q be the index for the households, q = 1, 2, 3,…., Q and let i be the index for the vehicle type alternatives. Let j be the index for the vehicle choice occasion j = 1, 2, …., J_q where J_q is the total number of choice occasions for a household q which is equal to N+2 (from RC data), plus the number of choice occasions where a replacement/addition decision was observed/reported (from SI data), plus up to eight choice occasions from the stated preference questionnaire (from SP data). With this notation, the vehicle type choice discrete component takes the following form:

$$u_{qij}^{*} = \beta' x_{qij} + \varepsilon_{qij} \tag{1}$$

u_{qij}^{*} is the latent utility that the qth household obtains from choosing alternative i at the jth choice occasion. x_{qij} is a column vector of known household attributes at choice occasion j (including household demographics and vehicle fleet characteristics before the jth choice occasion), β is the corresponding coefficient column vector of parameters to be estimated, and ε_{qij} is an idiosyncratic error term assumed to be independently and identically type-I extreme value distributed across alternatives, individuals, and choice occasions. Its scale parameter is normalized to one for revealed choice (RC) occasions and specified as $\dfrac{1}{\lambda}$ for the stated intention (SI) and stated preference (SP) choice occasions.

Then, the household q chooses alternative i at the jth choice occasion if the following condition holds:

$$u_{qij}^* > \max_{s=1,2,\ldots,I, s\neq i} u_{qsj}^* \tag{2}$$

The above condition can be written in the form of a series of binary choice formulations for each alternative i (*17*). Let R_{qij} be a dichotomous variable that takes the values 0 and 1, with R_{qij}=1 if the ith alternative is chosen by the qth household at the jth choice occasion, and R_{qij}=0 otherwise. Then, Equation (2) can be written as follows:

$$R_{qij} = 1 \text{ if } \beta' x_{qij} > v_{qij}, \ (i = 1, 2, \ldots, I) \tag{3}$$

$$\text{where } v_{qij} = \left\{ \max_{s=1,2,\ldots,I, s\neq i} u_{qsj}^* \right\} - \varepsilon_{qij} \tag{4}$$

The vehicle mileage component takes the form of a classical log-linear regression as follows:

$$m_{qij}^* = \alpha' z_{qij} + \eta_{qij}, \quad m_{qij} = 1 \quad \lfloor R_{qij} = 1 \rfloor m_{qij}^* \tag{5}$$

In the above equation, m_{qij}^* is a latent variable representing the logarithm of annual mileage for the vehicle type i if it had been chosen at the jth choice occasion. z_{qij} is the column vector of household attributes, α' is the corresponding column vector of parameter to be estimated, and η_{qij} is a normal error term assumed to be independent and identically distributed across households q and choice occasions j, and identically distributed across alternatives i ($\eta_{qij} \sim N[0, \sigma_\eta^2]$). Also, since the annual mileage is observed only for the chosen vehicle type at each choice occasion, any dependence between the m_{qij}^* terms across alternatives is not identified,

The two model components discussed above are brought together in the following equation system:

$$R_{qij} = 1 \text{ if } \beta' x_{qij} > v_{qij}, \ (i = 1, 2, \ldots, I) \ (j = 1, 2, \ldots, J)$$

$$m_{qij}^* = \alpha' z_{qij} + \eta_{qij}, \quad m_{qij} = 1 \quad \lfloor R_{qij} = 1 \rfloor m_{qij}^* \tag{6}$$

Copula based methods are used to determine the dependencies between the two stochastic terms v_{qij} and η_{qij} to account for common unobserved factors influencing vehicle type and usage decisions. In the copula method, the stochastic error terms are transformed into uniform distributions using their inverse cumulative distribution functions which are subsequently coupled into multivariate joint distributions using copulas (Eluru *et al.*, 2010). The expression for the log-likelihood is similar to the one in Eluru *et al.* (2010). Six different copulas were used in this study: (1) Gaussian copula, (2) Farlie-Gumbel-Morgenstern (FGM) copula, (3) Clayton, (4) Gumbel, (5) Frank, and (6) Joe copulas (Bhat and Eluru, 2009).

3.2.2 Vehicle Evolution Module

The vehicle selection module results are used even in the vehicle evolution module for predicting vehicle type and usage. In addition, a binary logit model form is used for modeling both the vehicle replacement and addition decisions (on an annual basis). Let *q* be the index for the households, *q* = 1, 2, 3,…., *Q*, let *i* be the index for the vehicle in the household and let *j* be the index for the vehicle replacement/addition occasion *j* = 1, 2, …., J_q where J_q is the total number of choice occasions for a household q which is equal to $\min\{t_{qi}, 5\}$, where t_{qi} is the number of years in which the household is planning to replace/add a vehicle *i*. For example, if a household with two vehicles plans to replace its first vehicle in two years, replace its second vehicle in five years, and add a vehicle in three years, then two choice occasions were created for the replacement decision of the first vehicle (0,1), five choice occasions for the replacement decision of the second vehicle (0,0,0,0,1), and three choice occasions for the addition decision (0,0,1), where 1 corresponds to an addition/replacement decision and 0 corresponds to a do-nothing option. With this notation, the vehicle evolution models take the following form:

$$l^*_{qij} = \gamma' w_{qij} + \vartheta_{qij}, \quad l_{qij} = 1 \text{ if } l^*_{qij} > 0\text{ ; } l_{qij} = 0 \text{ otherwise} \tag{7}$$

l^*_{qij} is the latent utility that the *q*th household obtains from choosing to replace/add vehicle *i* at the *j*th choice occasion. w_{qij} is a column vector of known household attributes at choice occasion j (including household demographics and vehicle fleet characteristics before the jth choice occasion), γ is the corresponding column vector of parameters to be estimated, and ϑ_{qij} is an idiosyncratic error term assumed to be independently and identically type-I extreme value distributed across alternatives, individuals, and choice occasions.

3.3 Data Description

The data for the current study is derived from the residential survey component of the California Vehicle Survey data collected in 2008-2009 by the California Energy Commission (CEC) to forecast vehicle fleet composition and fuel consumption in California. The survey included three components, which are briefly discussed in turn in the next three paragraphs.

The revealed choice (RC) component of the survey collected detailed information on the current household vehicle fleet and usage. This included information about the vehicle body type, make/model, vintage, and fuel type for each vehicle. In addition, the annual mileage that each vehicle is driven/utilized and the identity of the primary driver of each vehicle are also collected. The survey then included a set of questions to probe whether a household intended to replace an existing vehicle or acquire a net new additional vehicle in the fleet, and the characteristics of the vehicle(s) intended to be replaced or purchased (SI or stated intentions data). Essentially, the stated intention (SI) component of the survey gathered detailed information on replacement plans for each vehicle in the household fleet (over the next 25 years), and plans for adding net new vehicles (within the next five year period).

Finally, households that intended to purchase a vehicle either as a replacement or addition, and for whom there was adequate information on current revealed choices, were recruited for participation in a stated preference exercise (SP data). The SP exercises included several vehicle types and fuel technology options not currently available in the market, thus providing a rich data set for modeling vehicle transaction choices in a future context. The exercises involved the presentation of eight choice scenarios with four alternatives in each scenario. Attributes considered in describing each alternative included the vehicle type, size, fuel type, and vintage; a series of vehicle operating and acquisition cost variables; fuel availability, refueling time, and driving range; tax, toll, and parking incentives or credits; and vehicle performance (time to accelerate 0-60 mph).

The revealed choice (RC) and stated intentions (SI) data on current vehicle fleet composition and utilization was collected for a sample of 6577 households. Among these households, the stated preference (SP) component was administered to a sample of 3274 households who indicated that they would undertake at least one transaction in the future. The development of models for the vehicle simulator involved pooling the revealed choice (RC), stated intentions (SI) and stated preference (SP) components of the data, while pinning vehicle choice and usage behavior to current revealed choices.

The vehicle selection module estimation was undertaken using a random sample of 1165 respondent households with complete information. Care was taken to ensure that the distributions of vehicle types, fuel type and vintage in the estimation data set were the same as those in the original data set of 6577 observations. The discrete dependent variable in the vehicle selection module estimation is a combination of six vehicle body types (compact car, car, small cross utility vehicle, sport utility vehicle or SUV, van, and pick-up truck), seven fuel types (gasoline, flex fuel, plug-in hybrid, compressed natural gas (or CNG), diesel, hybrid electric, and fully electric), and five age categories (new, 1-2 years, 3-7 years, 8-12 years, and more than 12 years old). In addition, the no-vehicle choice category exists as well. Thus, there are a total of 211 alternatives in this choice process. The continuous dependent variable in the vehicle selection module estimation is the logarithm of the annual mileage traveled using each vehicle. The vehicle evolution component of the model system developed in this study includes the choice of replacement or addition of a vehicle. No information was collected on vehicle disposal plans and hence this choice dimension could not be considered using this data set. Of the 1165 household sample used for estimating the vehicle selection module, 915 households had complete information on vehicle transaction details (SI data). The replacement choice process is represented as an annual decision for each household, with replacement decisions beyond five years grouped into a single category of “five or more years”. Although the population is aged in the model estimation data set, many demographic changes are not taken into account (such as changes in number of workers, household income, household size, *etc*.) in the current effort; in ongoing work, the vehicle simulator described here is being integrated with a demographic evolution simulator to fully evolve households and their vehicle fleets over time.

3.4 Model Estimation Results

A sample of 1165 households with complete information provided the basis for estimating the model components. Descriptive statistics for this sample of households (as obtained from RC data) are shown in Table 3.1. Car, van, and SUV are the predominant vehicle types; annual mileage driven tends to be larger for larger vehicles than for cars, presumably because households use larger vehicles for longer trips. Less than two percent of households report having no vehicle. All of the other descriptive statistics show a reasonable distribution of attributes that makes the sample suitable for estimating choice models. The vehicle selection module includes the vehicle type choice model component (results are in Table 3.2) and the vehicle mileage component (results are in Table 3.3).

Table 3.1 Sample Characteristics

Variable	Sample Share (%)	Mean Mileage
Vehicle Type		
Compact Car	25.6	11894.36
Car	29.3	11887.08
Small Cross-utility Vehicle	4.8	11612.97
SUV	18.5	13099.24
Van	5.9	13019.13
Pickup	16.0	12310.61
Number of vehicles		
Zero	1.8	
One	28.4	
Two	50.0	
Three	14.2	
Four or more	5.6	
Number of adults		
One	18.5	
Two	64.3	
Three	10.7	
Four	4.9	
Five or more	1.5	
Number of workers		
Zero	18.3	
One	34.5	
Two	39.8	
Three	5.5	
Four or more	1.9	
Location		
Urban	48.2	
Suburban	47.8	
Rural	4.0	
Presence of senior adults	22.1	
Presence of children		
0-4 years	12.8	
5-11 years	14.9	
12 to 15 years	10.4	
Household Income		
<$20k	3.3	
Between $20 and $40K	13.1	
Between $40 and $60K	16.0	
Between $60K and 80K	18.3	
Between $80K and $100K	14.8	
Between $100K and $120K	10.8	
> $120K	23.7	
Educational Attainment		
High school	8.2	
College (with/without degree)	58.0	
Post Graduate	33.8	

Table 3.2 Vehicle fleet composition, utilization, and evolution simulator framework

Variable →	Constant	Generic Effects							
		Cost Variables				Vehicle Performance		Incentives	
		Purchase Price*10,000 ($)	Fuel cost per gallon ($)	Fuel cost per year /10,000 ($)	Maintenance cost per year/1000 ($)	Acceleration Time (0 to 60 mph)	Miles per Gallon /100	Car pooling	Free parking
No vehicle	--	--	--	--	--	--	--	--	--
Compact Car (CC)	-0.9371 (-5.95)								
Car	-1.3264 (-9.05)								
Small cross utility vehicle	-2.8986 (-14.28)								
SUV	-2.5797 (-15.38)								
Van	-3.5886 (-10.66)								
Pickup	-2.0160 (-11.89)								
Gasoline	-- --								
Flex Fuel	-6.2144 (-24.53)								
Plug-in Hybrid	-6.4622 (-16.20)								
CNG	-10.1330 (-12.47)	-0.6950 (-18.90)	-0.1469 (-1.86)	-4.7015 (-10.22)	-0.4843 (-2.35)	-0.0424 (-3.12)	4.8838 (13.59)	1.3079 (11.34)	1.4419 (12.19)
Diesel	-4.3522 (-18.67)								
Hybrid Electric (HE)	-4.1772 (-23.36)								
Fully Electric (FE)	-9.2407 (-12.46)								
New Car	-- --								
1 or 2 years	-1.9193 (-7.53)								
3 to 7 years	-1.3114 (-13.38)								
8 to 12 years	-3.1988 (-17.45)								
>12 years	-3.8380 (-14.78)								

TABLE 3.2 Estimates of the Vehicle Type Choice Component of Vehicle Selection Module (Continued)

Variable →	Generic Effects						Demographics					
	Incentives			Fuel Infrastructure/Vehicle Range					Household Income			
	$1,000 Tax credits	50% Reduced toll	$1,000 Vehicle price reduction	Fuel availability (1 in 50 stations)	Vehicle range (150 to 200 miles)	Vehicle range (>200 miles)	Number of male adults (>=16 years)	Number of female adults (>=16 years)	< $20K	($20K,$40K)	($40K,$60K)	($60K,$80K)
No vehicle	--	--	--	--	--	--	--	--	--	--	--	--
	--	--	--	--	--	--	--	--	--	--	--	--
CC				--	--	--	--	--	--	--	--	0.5159
				--	--	--	--	--	--	--	--	(4.67)
Car				--	--	--	0.4800	--	--	--	0.5436	1.1559
				--	--	--	(7.60)	--	--	--	(4.01)	(8.49)
SCU				--	--	--	--	--	--	--	--	0..9642
				--	--	--	--	--	--	--	--	(5.32)
SUV				--	--	--	0.3614	0.3614	--	--	0.3895	1.3496
				--	--	--	(7.85)	(7.85)	--	--	(2.31)	(8.82)
Van				--	--	--	0.5299	--	--	--	--	0.5645
				--	--	--	(4.02)	--	--	--	--	(3.56)
Pickup				--	--	--	0.6896	--	--	--	0.5322	0.8608
				--	--	--	(8.28)	--	--	--	(3.29)	(5.60)
Gasoline				--	--	--	--	--	--	--	--	--
				--	--	--	--	--	--	--	--	--
Flex Fuel				--	--	--	--	--	0.5122	0.5122	--	--
				--	--	--	--	--	(2.37)	(2.37)	--	--
Plug-in				--	--	--	-0.5595	--	-0.4032	--	--	--
				--	--	--	(-8.11)	--	(-1.52)	--	--	--
CNG	1.5135	1.1110	1.2653	-0.3278	4.6639	4.8415	--	--		--	--	--
	(17.74)	(9.83)	(10.53)	(-1.48)	(5.49)	(5.88)	--			--	--	--
Diesel				--	--	--	--	-0.4497	-0.9198	-0.9198	--	--
				--	--	--	--	(-4.06)	(-5.08)	(-5.08)	--	--
HE				--	--	--	-0.5595	--	--	--	--	0.3078
				--	--	--	(-8.11)	--	--	--	--	(2.88)
FE				-0.3278	4.6639	4.8415	--	0.4141	--	--	--	
				(-1.48)	(5.49)	(5.88)	--	(2.84)	--	--	--	--
New Car				--	--	--	--	--	--	--	--	--
				--	--	--	--	--	--			
1 or 2 years				--	--	--	--	--	--	0.5852	0.5852	0.5852
				--	--	--	--	--	--	(2.30)	(2.30)	(2.30)
3 to 7 years				--	--	--	--	--	--	0.5852	0.5852	0.5852
				--	--	--	--	--	--	(2.30)	(2.30)	(2.30)
8 to 12 yrs				--	--	--	--	--	--	0.9603	0.6543	--
				--	--	--	--	--	--	(6.71)	(4.35)	--
> 12 years				--	--	--	0.5111	--	--	0.9603	0.6543	--
				--	--	--	(3.17)	--	--	(6.71)	(4.35)	--

TABLE 3.2 Estimates of the Vehicle Type Choice Component of Vehicle Selection Module (Continued)

Variable →	Demographics										
	Household Income			Residential Location		Education Attainment		Presence of children			Presence of senior adults (>65 years)
	($80K,$100K)	($100K,$120K)	> $120K	Sub-urban	Rural	College	Post graduate	0 to 4 years	5 to 11 years	12 to 15 years	
No vehicle	--	--	--	--	--	--	--	--	--	--	--
	--	--	--	--	--	--	--	--	--	--	--
CC	0.5159	0.5159	0.8126	--	--	0.3971	0.5958	-0.2360	--	--	--
	(4.67)	(4.67)	(6.64)	--	--	(2.89)	(4.17)	(-1.86)	--	--	--
Car	1.1559	1.1559	1.6302	--	--	--	--	--	--	--	0.4286
	(8.49)	(8.49)	(11.19	--	--	--	--	--	--	--	(5.05)
SCU	0..9642	0..9642	1.8321	--	--	0.4175	--	--	-0.8584	--	--
	(5.32)	(5.32)	(9.56)	--	--	(3.05)	--	--	(-3.85)	--	--
SUV	1.3496	1.4079	1.8423	0.2403	--	0.1471	--	0.5392		--	--
	(8.82)	(8.04)	(11.28	(3.31)	--	(1.84)	--	(5.12)	--	--	--
Van	0.5645	0.5645	0.5645	--	--	0.6999	1.0881	1.1014	--	--	--
	(3.56)	(3.56)	(3.56)	--	--	(2.44)	(3.60)	(6.87)	--	--	--
Pickup	0.8608	0.7988	0.7988	0.5671	0.8937	--	-0.6031		--	--	--
	(5.60)	(4.89)	(4.89)	(5.98)	(3.96)	--	(-5.16)	--	--	--	--
Gasoline	--	--	--	--	--	--		--	--	--	--
	--	--	--	--	--	--		--	--	--	--
Flex Fuel	--	--	--	-0.2421	--	--	0.3105	--	--	--	-0.3524
	--	--	--	(-1.60)	--	--	(1.89)	--	--	--	(-1.88)
Plug-in	--	--	--	-0.3294	--	0.7447	1.4357	--	--	--	--
	--	--	--	(-2.97)	--	(2.63)	(4.78)	--	--	--	--
CNG	--	--	--	--	--	--	--	--	--	--	--
	--	--	--	--	--	--	--		--		--
Diesel	--	--	--	--	1.4089	-0.2817	--	-0.4497	--	0.3664	--
	--	--	--	--	(5.88)	(-2.52)	--	(-2.52)	--	(2.42)	--
HE	0.3078	0.3078	0.3078	-0.4084	0.6959	--	0.6418	--	--	--	0.3447
	(2.88)	(2.88)	(2.88)	(-4.71)	(2.24)	--	(6.70)	--	--	--	(3.49)
FE	--	--		-0.6467	--	1.5261	1.6286	0.7100	--	--	--
	--	--		(-4.04)	--	(2.56)	(2.69)	(3.66)	--	--	--
New Car	0.8084	0.8084	0.8084	--	--	0.2344	--	--	--	--	--
	(13.27)	(13.27)	(13.27	--	--	(3.44)	--	--	--	--	--
1 or 2	1.0202	1.0202	1.0202	--	--	--	-0.4539	--	--	--	--
	(3.90)	(3.90)	(3.90)	--	--	--	(-5.94)	--	--		
3 to 7	--	--	--	--	--	--	-0.4539	--	--	0.3980	-0.5208
	--	--	--	--	--	--	(-5.94)	--	--	(4.74)	(-6.26)
8 to 12 yrs	--	--	-	--	--	--	-0.4539	--	-0.5472	0.3980	-0.5208
	--	--	(-3.83)	--	--	--	(-5.94)	--	(-3.10)	(4.74)	(-6.26)
>12 years	--	--	-	--	--	--	-0.4539	--	-0.5472	0.3980	-0.5208
	--	--	(-3.83)	--	--	--	(-5.94)	--	(-3.10)	(4.74)	(-6.26)

TABLE 3.2 Estimates of the Vehicle Type Choice Component of Vehicle Selection Module (Continued)

Variable→	Demographics					Existing Fleet Characteristics					
		Number of workers									
	Caucasian	# full time workers	# part time workers	# full time workers from home	# part time workers from home	Presence of CC	Presence of Car	Presence of SCU	Presence of SUV	Presence of Van	Presence of pickup
No vehicle	--	--	--	--	--	--	--	--	--	--	--
	--	--	--	--	--	--	--	--	--	--	--
CC	0.1266	--	--	--	0.2752	-1.9803	-2.0374	-0.3408	-2.0862	-0.5126	-0.8680
	(1.80)	--	--	--	(1.63)	(-13.15)	(-11.56)	(-1.95)	(-10.38)	(-2.68)	(-4.41)
Car	0.1748	-0.0933	0.3942	--	--		-2.2192	--	-2.0862	-0.2859	-0.7981
	(2.53)	(-1.97)	(6.45)	--	--		(-12.72)	--	(-10.38)	(-1.73)	(-4.36)
SCU	--	--	--	--	--	-0.8672	-1.1525	--	-1.2043		-0.9770
	--	--	--	--	--	(-4.11)	(-5.36)	--	(-4.92)		(-3.67)
SUV	--	--	--	0.3456	0.3316	-1.6154	-1.6188	--	-1.8460	-0.2859	-0.6969
	--	--	--	(1.68)	(1.85)	(-9.91)	(-9.32)	--	(-9.17)	(-1.73)	(-3.67)
Van	--	--	--	0.3456	0.6416	-1.3314	-1.2999	--	-1.8460	-1.1981	-0.5083
	--	--	--	(1.68)	(2.43)	(-6.54)	(-5.79)	--	(-9.17)	(-3.63)	(-2.13)
Pickup	--	--	0.2404	--	--	-1.6384	-1.6229	--	-1.8460		-1.7183
	--	--	(2.87)	--	--	(-9.03)	(-8.42)	--	(-9.17)		(-8.11)
Gasoline	--	--	--	--	--			-0.5164		-1.1119	--
	--	--	--	--	--			(-4.32)		(-10.27)	--
Flex Fuel	--	--	--	-0.9011	--	1.9187	1.3517	--	2.0346	--	0.8025
	--	--	--	(-1.63)	--	(10.17)	(7.31)	--	(14.75)	--	(3.48)
Plug-in	-0.1816	--	--	-0.7593	--	1.8859	1.2428	--	2.0346	--	0.6614
	(-2.16)	--	--	(-2.96)	--	(11.73)	(6.93)	--	(14.75)	--	(3.69)
CNG	-0.1816	--	--	-1.5793	1.0713	0.8919	0.9285	--	1.2867	--	--
	(-2.16)	--	--	(-1.67)	(2.86)	(3.33)	(3.56)	--	(4.46)	--	--
Diesel	--	0.1132	--	--	--	1.8670	1.4401	--	1.5186	--	0.5451
	--	(1.52)	--	--	--	(11.78)	(8.41)	--	(8.05)	--	(3.28)
Hybrid	--	--	--	--	0.5104	1.1027	0.8652	-0.5752	1.5457	-0.5590	0.4686
	--	--	--	--	(2.34)	(8.21)	(7.16)	(-2.61)	(11.64)	(-2.86)	(3.12)
Fully	-0.1816	--	--	--		0.5449	0.6123	--	1.0249	--	--
	(-2.16)	--	--	--	--	(2.61	(3.19)	--	(5.03)	--	--
New Car	--	--	--	0.4248	--	-1.0488	-1.1421	-0.9662	-1.1690	-1.2475	-0.9937
	--	--	--	(4.05)	--	(-7.15)	(-7.08)	(-5.17)	(-5.96)	(-6.92)	(-5.48)
1 or 2 years	--	0.1556	0.2131	--	--	-0.6136	-0.6546	-0.9662	-0.7563	-0.8028	-0.5891
	--	(3.12)	(3.32)	--	--	(-4.43)	(-4.09)	(-5.17)	(-3.81)	(-4.16)	(-3.24)
3 to 7 years	--	0.2518	0.3530	--	--	-0.103	-0.6546	-0.7738	-0.7563	-0.8028	-0.5891
	--	(5.87)	(5.64)	--	--	(-5.92)	(-4.09)	(-3.62)	(-3.81)	(-4.16)	(-3.24)
8 to 12	0.4773	0.2673	0.3782	--	--	--	--	--	--	--	--
	(4.42)	(4.35)	(4.26)	--	--	--	--	--	--	--	--
>12 years	0.4773	0.2673	0.3782	--	--	--	--	--	--	--	--
	(4.42)	(4.35)	(4.26)	--	--	--	--	--	--	--	--

TABLE 3.2 Estimates of the Vehicle Type Choice Component of Vehicle Selection Module (Continued)

Variables →	Replaced Vehicle Characteristics						
	Compact Car	**Car**	**SCU**	**SUV**	**Van**	**Pickup**	**Gasoline**
No vehicle	--	--	--	--	--	--	--
CC	0.5665	-1.5864	-0.9648	-1.3750	--	-2.1573	1.5717
	(2.17)	(-8.82)	(-2.60)	(-4.44)	--	(-6.32)	(6.06)
Car	--	1.9106		0.9306	1.1680	-0.7985	--
	--	(12.55)		(4.41)	(3.86)	(-2.90)	--
SCU	--	--	2.5700	--	--	--	0.5343
	--	--	(9.41)	--	--	--	(3.20)
SUV	--	--	--	2.6388	1.6229	--	--
	--	--	--	(12.72)	(5.97)	--	--
Van	--	--	--		4.7040	--	--
	--	--	--		(13.34)	--	--
Pickup	--	-0.4319	--	1.3940	--	4.3382	-0.8290
	--	(-1.64)	--	(5.06)	--	(15.92)	(-4.47)
Gasoline	-0.4069	--	-0.6777	--	--	-1.14	--
	(-3.46)	--	(-3.11)	--	--	(-5.24)	--
Flex Fuel	--	--	--	--	--	-0.8779	0.7836
	--	--	--	--	--	(-2.69)	(3.61)
Plug-in Hybrid	--	0.5869	--	--	0.8307	-0.9392	0.8037
	--	(2.79)	--	--	(2.87)	(-2.90)	(3.77)
CNG	--	--	--	--	--	--	--
	--	--	--	--	--	--	--
Diesel	--	0.7886	--	1.0441	--	0.7583	-0.6766
	--	(3.63)	--	(4.27)	--	(2.54)	(-3.82)
Hybrid Electric	--	--	--	--	--	-1.6336	1.5442
	--	--	--	--	--	(-5.89)	(12.07)
Fully Electric	--	--	--	--	--	--	-0.5583
	--	--	--	--	--	--	(-2.32)
New Car	-0.1958	--	--	1.7986	--	0.4506	3.3215
	(-1.61)	--	--	(2.84)	--	(2.82)	(8.10)
1 or 2 years	--	--	--	1.7986	--	--	3.3215
	--	--	--	(2.84)	--	--	(8.10)
3 to 7 years	--	--	--	1.7986	--	--	3.3215
	--	--	--	(2.84)	--	--	(8.10)
8 to 12 years	--	--	--	--	--	--	2.0138
	--	--	--	--	--	--	(4.66)
	--	--	--	--	--	--	--

Table 3.3 Estimates of the Vehicle Usage Component of Vehicle Selection Module

Variable	Parameter	t-stat
Constant	8.4682	128.77
HH Income		
Above $80K	0.0401	2.25
Presence of children		
Under 4 years	0.0398	1.58
Location of HH		
Sub-urban	0.1074	6.61
Presence of senior adults (age>65 years)	-0.1281	-5.97
Number of vehicles		
Two	-0.0662	-2.71
Three	-0.1667	-5.56
Four	-0.2524	-6.21
Number of workers	0.0763	6.83
Mean distance to work /10 (miles)	0.091	12.67
Vehicle Characteristics		
Car	0.0446	1.85
Small cross utility vehicle	-0.1329	-3.01
SUV or Van	0.0767	2.93
8 to 12 years old	-0.4298	-8.09
More than 12 years old	-0.7189	-12.87
Standard error of the estimate	0.7476	42.42
Scale Parameter (λ)	0.5538	23.91*
Copula Dependency Parameter (θ)	-3.4097	-9.38

* t-statistic computed against a value of 1

3.4.1 Vehicle Selection Module

For the vehicle type component, the overall utility of a vehicle type is considered as the sum of independent utility components for the body type, fuel type, and vintage of the vehicles. While certain interaction effects were also considered, such effects were generally not statistically significant. Thus, Table 3.2 presents the effects of variables in three row panels: the first row panel corresponds to body types (including the "no vehicle" option), the second to fuel types, and the third to vehicle vintage. The results offer behaviorally intuitive interpretations. Strictly speaking, the constants (first column of Table 3.2) cannot be directly compared across the body types because of the presence of several continuous variables in the model specification, but the magnitudes of the constants on the different body types suggest a greater preference to own a compact car or a car compared to other vehicle types. In the second row panel, similarly, gasoline fuel vehicles are the most preferred, while compressed natural gas (CNG) and fully electric vehicles are the least preferred. The final row panel suggests, as expected, that households have a strong preference for newer cars.

A range of policy-sensitive variables were included in the model, as shown in Table 3.2. These are all estimated as generic effects (that is, a single effect is estimated for each variable across all alternatives as indicated by the dotted lines separating the three panels in Table 3.2). All of the cost-related variables (purchase price, fuel cost per gallon, fuel cost per year/$10000, and maintenance cost per year/$1000) have negative coefficients indicating that as cost increases, the preference for a vehicle type decreases. Two vehicle performance variables were considered. The time to accelerate from 0 to 60 mph has a negative impact on the utility of an alternative, indicating that, in general, vehicles with more powerful engines are preferred. Similarly, fuel efficiency (measured in miles per gallon) also has a positive impact on utility. Interestingly, policy variables that offered incentives such as car pooling, free parking, $1000 tax credit, 50 percent reduction in tolls, and $1000 off the purchase price all have similar magnitudes of effects on enhancing the utility of various alternatives. In other words, one policy incentive did not clearly outshine the others in terms of influencing vehicle type choice. But, all these policy variables are statistically significant in the final model.

In the category of fuel infrastructure and vehicle range, for CNG and electric vehicles, the greater availability of refueling stations positively affects vehicle type choice (note the negative sign on the “fuel available – 1 in 50 stations” variable in Table 3.2; the base for introducing this variable was “fuel available – 1 in 20 stations”). Refueling time, however, did not turn out to be statistically significant. Also, for CNG and electric vehicles, those with medium (150-200 miles) and high (>200 miles) driving ranges are preferred over those with lower ranges.

As expected, a range of household socio-economic and demographic variables significantly affects vehicle type choice. Households with more male adults have a stronger preference (relative to households with fewer males) for larger vehicles as opposed to compact cars and small cross utility vehicles, and were more likely to own older (>12 years) vehicles (an adult is defined as an individual over 15 years of age). Interestingly, these households have a lower preference for plug-in hybrid and hybrid electric vehicles than households with fewer males. On the other hand, households with more female adults have a higher propensity (than households with few female adults) to own sports utility vehicles (SUVs) and move toward owning fully electric vehicles, while also shying away from diesel-powered vehicles.

As household income increases, the inclination to acquire older vehicles decreases. Higher income households are likely to be able to afford newer vehicles and have a preference to do so. Also, higher income households show a preference for a mix of vehicle body types including both small and large vehicles, suggesting that these households are able to afford a mix of vehicle body types for different types of trips. Households located in suburban regions are more inclined to own regular gasoline or diesel or CNG fueled sports utility and/or pick-up vehicles, while households in rural areas are more likely to own pick-up vehicles and diesel/hybrid fueled vehicles (the base category was households residing in urban regions). Those with a higher education level tend to have a preference for newer vehicles and alternative fuel vehicles. It is possible that these individuals are more environmentally sensitive, leading to their preference for less polluting vehicles (the education level of high school or below was the base category for introducing education effects). Households with younger children tend to prefer larger vehicles, consistent with the notion that families probably like the room offered by such vehicles. Households with older children have a preference for acquiring older vehicles, perhaps because parents purchase older vehicles for teenagers when they first begin driving. On the other hand, households with senior adults (>65 years of age) prefer newer vehicles, possibly because these households want trustworthy cars that are perceived to be safe.

A set of findings hard to explain is that Caucasian households are more likely to prefer cars over larger vehicles, older vehicles over newer vehicles, and traditional fuel vehicles over alternative fuel vehicles. It is not immediately clear why these preferences exist for this group in comparison to other groups. Similarly, it is not readily apparent why households with more full-time and part-time workers with a work location outside home should prefer older cars relative to new cars, while households with several full-time workers working from home would have a propensity to own new cars. Finally, households with several employed individuals working from home are more likely to own SUVs and vans.

The existing household vehicle fleet has a significant impact on vehicle type choice/selection. Among the many effects of existing household fleet, the one that particularly stands out is that households prefer less any vehicle body type that already exists in their fleet. With respect to replacement (last page of Table 3.2), there are several tendencies, but an overarching result is that households are more prone to replace a vehicle in the fleet with the same body type of vehicle. If the replaced vehicle is a compact car, it is likely to be replaced with a non-gasoline fueled vehicle but also not the newest of vehicles (possibly because current compact car owners are more environmentally conscious but also cost-conscious, which leads them to seek "green" vehicles but not the newest vehicles). A car is unlikely to be replaced with a pick-up. Also, in general, any non-compact car is unlikely to be replaced with a compact car.

When the replaced vehicle is a SUV, households tend to replace it with a diesel-powered engine, and with a newer vehicle rather than an older one. Households which replace a gasoline fuel vehicle are more likely to replace it with an alternative fuel vehicle rather than a diesel fuel vehicle. This suggests that households looking to replace an existing gasoline vehicle are likely to consider newer alternative fuel vehicles; public policies aimed at offering incentives may provide the needed impetus to move in the direction of a greener fleet.

The vehicle usage (mileage) model component in Table 3.3 also yielded largely intuitive results. Households with higher incomes are associated with higher travel mileage, consistent with the notion of more financial freedom to engage in out-of-home discretionary pursuits. Households with small children tend to have larger mileage, perhaps because these households have more errands to run and serve-child trips that accumulate miles. Households in suburban regions also travel more than other households, possibly because suburban locations are more auto-oriented. Households with senior adults greater than 65 years of age tend to have lower mileage, presumably because these households consist of retired individuals living in empty nests. Households with more vehicles have lower mileage on a per vehicle basis, a manifestation of the ability to divide total household travel among multiple vehicles. Households with more workers have larger mileage, presumably due to greater levels of work travel. Similarly, households in which individuals are farther from their work places accumulate more mileage on their vehicles. Higher mileage values are associated with cars and larger vehicles such as SUV and van, but lower mileage values are associated with smaller cross utility vehicles and older vehicles.

The vehicle selection module of Figure 3.1 was estimated by pooling RC, SI and SP data. In such pooled estimations, one is often concerned with the possibility that the choice process exhibited in the RC data is different from that exhibited in the SI and SP data. For this reason, a scale parameter was estimated in the vehicle type choice – usage model to adjust model parameters in the joint RP-SI-SP model system. The RP to SI-SP scale parameter (λ) was estimated to be 0.5538 with a t-statistic of 23.91 (against a value of 1 which corresponds to the case when the variance of unobserved factors in the RP and SI-SP contexts are equal). This scale parameter is significantly smaller than unity, indicating that the error variance in the SI-SP choice context is higher than in the RP choice context (see Borjesson (*19*) for similar result).

Among all the copula structures considered, the Frank copula model offered the best statistical fit based on the Bayesian Information Criterion (BIC) (*20*). The corresponding copula dependency parameter (θ) was estimated to be equal to -3.4097 with a t-statistic of -9.38. This shows that there is significant dependency between the vehicle type choice and usage dimensions. The Kendall's measure (τ) which is similar to the standard correlation coefficient was computed using the expression:

$$\tau = 1 - \frac{4}{\theta}\left[1 - \frac{1}{\theta}\left[\int_{t=0}^{\theta} \frac{t}{e^t - 1} dt\right]\right]$$

The value of τ was found to be -0.3411. The error term ν_{qij} enters Equation (3) with a negative sign. Thus, a negative sign on the Kendall's measure indicates that the unobserved factors that increase the propensity to choose a certain vehicle type also increase the propensity to accumulate more mileage on that vehicle.

In terms of data fit, the log-likelihood value at convergence of an independent model that models vehicle type choice and usage separately was -29382.7. The Frank copula model, which offered the best statistical fit among all the joint copula model structures, had a log-likelihood value of -29187.20 The improvement in fit, relative to the independent model, is readily apparent and is highly statistically significant. To demonstrate that this improvement is not simply an artifact of overfitting, an additional evaluation exercise was undertaken to test the comparative ability of the independent and joint models to replicate vehicle fleet composition choices in a random hold-out sample of 500 households not included in the estimation sample (see Table 3.4). The predicted log-likelihood function values of the independent and copula-based joint models were compared for different segments of the hold-out sample. The overall predictive log-likelihood ratio test values for comparing the copula based joint model with the independent model indicate that the copula based joint model is significantly better than the independent model in all cases from a statistical standpoint, except for households with no vehicles and households that have four or more workers where there is no appreciable difference in predictive power between the two models. The results clearly demonstrate the superiority of the joint model in predicting vehicle fleet composition and utilization, relative to the independent model.

Table 3.4 Disaggregate Measures of Fit for the Validation Sample

Sample details	Number of households	Independent model predictive likelihood	Copula based joint model predictive likelihood	Predictive likelihood ratio test ($\chi^2_{1,0.05} = 3.84$)
Full validation sample	500	-14189.96	-14084.80	208.29
Number of vehicles				
Zero	6	-157.011	-156.08	1.86
One	152	-3030.74	-3013.22	35.04
Two	225	-6337.90	-6298.90	77.99
Three	89	-3292.88	-3256.84	72.09
Four or more	28	-1370.43	-1359.78	21.30
Number of workers				
Zero	90	-2123.99	-2116.89	14.20
One	171	-4513.83	-4484.28	59.08
Two	196	-5857.35	-5806.80	101.09
Three	37	-1380.86	-1365.08	31.57
Four or more	6	-312.93	-311.77	2.32
Highest Educational Attainment				
High school	43	-1117.53	-1108.82	20.68
College (With/without degree)	271	-7768.68	-7726.33	100.78
Post Graduate	186	-5302.75	-5271.41	86.83
Presence of children				
0-4 years	57	-1679.78	-1661.28	37.00
5-11 years	74	-2197.82	-2179.51	36.63
12-15 years	58	-1917.09	-1891.06	52.06
Presence of senior adults (Age≥65 years)	113	-2902.10	-2890.35	23.51
Region				
Urban	241	-6704.93	-6652.75	104.36
Sub-urban	235	-6785.54	-6740.34	90.40
Rural	24	-698.49	-691.72	13.53

3.4.2 Vehicle Evolution Models

The vehicle evolution model component consists of an annual replacement decision model and an addition decision model. Estimation results for the replacement and addition models are presented in Tables 3.5 and 3.6 respectively, and are discussed here.

Table 3.5 Replacement Decision of Evolution Module: Binary Logit Model

Variable	Parameter	t statistic
Constant	-1.9667	-8.84
Race of household (other race is base)		
Caucasian	0.1108	1.59
Hispanic	0.7353	1.43
Household Income (Base is below $60,000)		
Between $60,000 and $100,000	0.1065	1.26
Above $120,000	0.1689	1.76
Presence of children		
5 to 11 years	-0.1736	-1.79
12 to 15 years	0.4677	3.20
Characteristics of vehicle getting replaced		
Small cross utility vehicle	-0.4269	-2.21
SUV	-0.2567	-2.57
SUV*Large Household	-0.4565	-2.23
Van	-0.2168	-1.55
Pickup	-0.1997	-1.92
1-3 years old	0.1432	1.40
3-7 years old	0.3125	3.23
8-12 years old	0.6889	4.18
More than 12 years old	0.548	3.01
Gasoline Fueled	0.3529	1.71
Number of years since acquired (Base is 5 or more years)		
1 year	-1.8907	-4.81
2 years	-1.1948	-5.96
3 or 4 years	-0.8159	-8.02
Number of years since a vehicle has been replaced	0.5908	14.23
Number of years since a vehicle has been added	0.2910	3.31
Log Likelihood	-2675.62	
Log Likelihood at constants	-2892.99	

Table 3.6 Addition Decision of Evolution Module: Binary Logit Model

Variable	Parameter	t statistic
Constant	-3.7901	-5.60
Race of the household (other race is base)		
Caucasian	-0.4064	-1.77
Hispanic	-9.576	-9.49
Number of adults	0.8129	5.14
Large Household (size >=5)	0.7117	2.16
Household Income (Base is above $20,000)		
Between $20,000	1.4209	2.96
Presence of children 12 to 15 years	1.2988	4.48
Presence of senior adults (age >65 years)	-1.8651	-3.36
Region (Base is urban and sub-urban)		
Rural	0.9864	2.07
Household Vehicle Fleet Characteristics		
Number of compact cars	-0.7671	-3.16
Number of cars	-0.4622	-2.01
Number of SUVs	-0.2942	-1.57
Number of Pickup trucks	-0.5665	-2.28
Number of years since a vehicle has been replaced (Base is four or more years)		
Same year	-1.0295	-1.62
One to three years	-0.8189	-1.28
Log Likelihood	-428.88	
Log Likelihood at constants	-506.45	

The replacement model is a binary logit model that was found to offer plausible behavioral findings. The constant is significantly negative suggesting that households have a baseline preference to not replace their vehicles from one year to the next; this is consistent with the notion that vehicle transactions are infrequent events often spaced years apart. Caucasian and Hispanic households are more likely to replace a vehicle than households of other races. As expected, higher income households are more likely to replace a vehicle, while those with young children are less inclined to replace a vehicle. It is possible that households with young children are dealing with new expenses and do not feel the need to replace a vehicle. Households with older children are more likely to replace a vehicle, possibly because their fleet is getting old or because they are getting ready for the day when one or more children begins to drive. Small cross-utility vehicles are the least likely to be replaced; van, SUV, and pick-up truck are also not very likely to be replaced, and this reluctance to replace is particularly so for SUVs in large households. Among all body types, compact cars and cars (the base body type categories) are the most likely to be replaced. Older vehicles are more likely to be replaced than newer ones, although the coefficient for the 12 years or older category is less positive than for the 8-12 year old category. It is possible that vehicles 12 years or older have either been maintained very well, had parts replaced, or simply hold an emotional attachment that reduce the likelihood of replacement compared to the 8-12 year old category. Gasoline fuel vehicles are the most likely vehicle fuel type to be replaced, a finding consistent with the fact that gasoline vehicles are the predominant vehicle type in the population. Vehicles which are held for five or more years are most likely to be replaced, and the propensity to replace reduces (increases) as the duration of ownership decreases (increases). Finally, as expected, the results suggest important interdependencies in the transaction history. That is, the longer the duration (*i.e.,* number of years) since any other vehicle in the household has been replaced or a vehicle has been added, the more likely that the household will replace a vehicle it currently holds (note that these variables are created based on the planned replacement or addition of vehicles, as obtained from the stated intentions data).

The vehicle addition model is also a binary logit model. Hispanic households are found to be the least likely to add a vehicle. Caucasians are found to be the second least likely to add a vehicle. Households with more adults and a larger number of persons are more likely to add a new vehicle to their fleet. Lower income households are found to be more likely to add a vehicle in comparison to other higher income categories. It is possible that lower income households do not currently have the desired number of vehicles and hence desire to add a net additional vehicle to the fleet. Higher income households probably have the desired number of vehicles and so, rather than add a net additional vehicle, merely wish to replace an existing vehicle over time. Households with senior adults are less inclined to add a vehicle, while households with children aged 12-15 years are more likely to add a vehicle presumably because they are getting to acquire a vehicle for the new driver in the household. Households in rural regions appear more likely to add a vehicle. The larger the current vehicle fleet size, the less likely it is for a household to add a net additional vehicle. This is true across all vehicle type categories. Finally, the results indicate that it is less likely for a household to add a vehicle if a vehicle has been replaced recently. It was not possible to include the effect of recent vehicle additions on the decision to add a vehicle because only eight households in the data indicated that they would add two new vehicles within the next five years.

The log-likelihood values at convergence of the replacement and addition models are -2675.62 and -428.88 respectively. The corresponding values for the "constant only" models are -2892.99 and -506.45 respectively. Clearly, one can reject the null hypothesis that none of the exogenous variables provide any value to predicting decision to replace/add a vehicle at any reasonable level of significance.

3.5 Conclusions

The modeling and analysis of household vehicle ownership and utilization by type of vehicle has gained added importance in recent years in the face of rising concerns about global energy sustainability, greenhouse gas (GHG) emissions, and community livability in urban areas around the world. Households may choose to own and drive (utilize) a variety of different vehicle types and the ability to accurately forecast these choice dimensions is undoubtedly of much interest in the current planning context that is dominated by efforts on the part of planners and policy makers to minimize the adverse impacts of automobile use on the environment.

This study presents the design and formulation of a comprehensive vehicle fleet composition and evolution simulator that is capable of simulating household vehicle ownership and utilization decisions over time. The simulation framework consists of two main modules – one module that models the current (baseline) fleet composition and utilization for a household and another module that evolves the baseline fleet over time by considering the acquisition, replacement, and disposal processes that households may undertake as they turnover their fleet.

One of the major impediments thus far to the development of such a vehicle fleet evolution simulation system has been the availability of longitudinal data on the dynamics of household vehicle ownership and utilization by type of vehicle. This issue is overcome in this study through the use of a large sample data set collected as part of a survey undertaken by the California Energy Commission in California. The survey includes a revealed choice (RC) component that captures information about current vehicle fleet information for the respondent households, a stated intentions (SI) component that captures information on the plans of respondent households to replace existing household vehicles or add net additional vehicles to the fleet (and the timing of such potential transactions), and a stated preference (SP) component that captures information on the vehicle type likely to be chosen by households when faced with a set of hypothetical choice scenarios. Data from these three survey components are pooled together to obtain a rich data set that can be used to model the full range of vehicle ownership and transactions decisions of households.

The joint modeling framework is applied to predict vehicular choices for a random holdout sample of households and shown to perform substantially better than an independent set of model components that ignore common unobserved factors that impact both vehicle fleet composition and utilization. The approach presented in this study offers the ability to generate vehicle fleet composition and usage measures that serve as critical inputs to emissions forecasting models. The novelty of the approach is that it accommodates all of the dimensions characterizing vehicle fleet/usage decisions, as well as all of the dimensions of vehicle transactions (*i.e.*, fleet evolution) over time. The resulting model can be used in a microsimulation-based forecasting model system to obtain the fleet composition for a future year and/or examine the effects of a host of policy variables aimed at promoting vehicle mix/usage patterns that reduce GHG emissions and fuel consumption.

CHAPTER 4: Vehicle Fleet Forecasting

The vehicle fleet simulator developed in this study is used to simulate the effects of a multitude of policy actions, and analyze the predicted vehicle fleet composition and usage results through sensitivity tests. That is, the vehicle composition and usage forecasts as well the associated fuel consumption and GHG emissions are made under several alternative technology and policy scenarios for future years. This chapter documents the results of this analysis. Specifically, Section 4.1 provides a brief overview of the vehicle fleet simulator and Section 4.2 describes the forecasting procedure including the population evolution system and the vehicle fleet evolution system. Section 4.3 presents the forecasting results and Section 4.4 concludes the work by summarizing key findings.

4.1 Methodology

4.1.1 Overview of the Vehicle Fleet Simulator

As discussed earlier in Chapter 3, all the models in the vehicle fleet simulator are estimated using a unique dataset that includes comprehensive information on vehicle ownership and usage decisions of households, including current fleet composition, potential future fleet composition, and vehicle evolution plans. The vehicle fleet simulator incorporates innovative methodological approaches to address the problem of multiple vehicle holdings and use, as well as to deal with the gamut of vehicle evolution decisions, all in a comprehensive and implementable forecasting framework. Specifically, the simulator encompasses state-of-the-art household vehicle type choice, usage, and evolution models estimated using a unique 2008-2009 vehicle survey data set collected from 6577 households in the State of California by Resource Systems Group, Inc. (RSG) for the California Energy Commission (CEC). The survey has three components - (1) a revealed choice (RC) component, which collected information about current vehicle holdings and usage, (2) a stated intentions (SI) component, which collected information on replacement plans for existing vehicles and vehicle addition plans, and (3) a stated preference (SP) component, which collected information about vehicle purchase decisions under hypothetical scenarios.

The vehicle fleet simulator consists of two principal components- (1) The vehicle selection module, and (2) The vehicle evolution module. Each vehicle type alternative in the vehicle selection module is defined as a combination of six vehicle body types (compact car, car, small cross utility vehicle, sport utility vehicle or SUV, van, and pick-up truck), seven fuel types (gasoline, flex fuel, plug-in hybrid, compressed natural gas (or CNG), diesel, hybrid electric, and fully electric), and five age categories (new, 1-2 years, 3-7 years, 8-12 years, and more than 12 years old). Thus, there are a total of 211 vehicle type alternatives including the alternative of no-vehicle. The model system accommodates multiple vehicle ownership and usage dimensions by assuming that vehicle fleet and usage decisions are determined through a series of unobserved (to the analyst) repeated discrete-continuous choice occasions. This framework mimics the dynamics in the vehicle acquiring process by accommodating the impacts of the types of vehicles already owned on the type of vehicle that may be purchased in a subsequent purchase decision. The number of choice occasions in such a "vertical" choice behavior is linked to the number of adults in the household. In particular, since the number of vehicles is almost never greater than the number of adults in the household plus two in the data, the number of choice occasions is set to be equal to the number of adults plus two. At each choice occasion, the household may choose not to purchase a vehicle or to acquire a vehicle of a certain type. However, the choice of vehicle ownership, vehicle type and vehicle utilization are likely to be multiple dimensions of a single choice bundle at each choice occasion. This joint nature of decisions is recognized at each choice occasion by using a copula-based joint discrete-continuous framework. Also, SP and SI choice behavior is pinned to revealed choice behavior by adopting a combined revealed preference-stated intention-stated preference estimation technique of including a scale parameter differential between the RC and the SP and SI processes. In the framework, the decision of the number of vehicles owned by the household is endogenously, even if implicitly, determined as the sum of those choice occasions when the household selects a certain vehicle type. Overall, the vehicle selection module jointly models all base year vehicle fleet characteristics in a unifying framework.

In the vehicle evolution module, the number of choice occasions for evolving the vehicle fleet each year is set to the current vehicle fleet plus one (this assumes that households do not add more than one vehicle to their current fleet, after considering replacements; however, the model structure easily handles any number of additional vehicles by increasing the number of choice occasions to number of vehicles plus "x", where "x" is appropriately chosen depending upon the empirical context). For the choice occasions corresponding to an existing vehicle, the household has three options: (1) Keep the vehicle, (2) Dispose the vehicle, and (3) Replace the vehicle (and vehicle type and usage of the replacement vehicle). For the choice corresponding to the last choice occasion, the alternatives are "not to add a vehicle" or "to add a vehicle" along with vehicle type and usage of the added vehicle. All the models in the evolution module are binary logit models that consider dependency between transaction decisions by including the number of years since an earlier transaction decision as an explanatory variable in the utility specification. The vehicle type and usage of all the replacement /added vehicles are determined using the vehicle type choice model in the vehicle selection module. The vehicle type choice model includes existing vehicle fleet characteristics and the replaced vehicle characteristics as explanatory variables in the utility specification of vehicle type alternatives. This captures dependencies between future vehicle type choices (during evolution) and vehicles already owned and the vehicle getting replaced.

4.2 Forecasting Process

4.2.1 Population Evolution

We use the entire estimation sample of 6577 households to undertake the forecasting simulations, because these households are sampled to be representative of the population in the State of California. For predicting future vehicle holdings and usage, this base year population is evolved into the future using a suite of models. The evolution framework used in this study is borrowed from a comprehensive evolution module of the activity-based travel demand model "SimAGENT" currently under development for the Southern California region known as CEMSELTS developed by Bhat and colleagues at the University of Texas at Austin (see Pendyala *et al.*, 2012 and Eluru *et al.*, 2008).

Several demographic processes including ageing, death, birth, immigration, move-out of young adults, marriages, and divorces are modeled in the framework. Figure 4.1 shows the population evolution framework adopted in this study. While a more comprehensive evolution framework may be adopted, for the purpose of this study we only evolve key demographic characteristics that influence the vehicle type choices of the household (*i.e.,* appear as explanatory variables in the vehicle type choice or vehicle evolution choice models). First, we model the immigration of new households into the study region. While the actual immigration process involves several other processes including immigration of international population into the study region, domestic immigration, and emigration of resident population out of study region, for the purpose of this study, we use net immigration rate after accounting for all these three processes. The latest estimate on the immigration rates available from the US Census Bureau is 1.8 per thousand average population (US Census Bureau, 2009). The characteristics of the immigrating households are determined by randomly assigning the characteristics of one of the households in the study region. The residential location of the immigrant households (urban, sub-urban, or rural) is obtained using probability rates from the original survey data.

Figure 4.2 shows all the components of the individual-level evolution and choice models. Historical mortality rates by age, gender, and ethnicity from 1995 to 2009 provided by the State of California's Department of Public Health formed the basis for obtaining the mortality rates for the future years (CDPH, 2010). The National Centre for Health Statistics projected national mortality rates for 2010-2050 (http://www.nfda.org/media-center/statisticsreports.html). We calculated the California mortality rates for the future years by adjusting the mortality rates of California in proportion to the national mortality rates. For birth rates, we used the latest rates by age and ethnicity of the mother as projected by the Department of Finance in 2011 using the most recent 2010 Census data (DOF, 2011). The education attainment (less than high school, high school, bachelor's degree, or post graduate degree) and employment characteristics (full time *versus* part time, and distance to work place) of new individuals was determined using probability rates observed in the original survey data.

The marriage rate (17.5 marriages per 1000 women) and divorce rate (8.9 divorces per 1000 women) for the year 2009 for the State of California are obtained from the American Community Survey data (ACS, 2011). We assumed the same marriage and divorce rates for all future years. The education and work characteristics of the spouse were obtained using probability rates observed among married couples in the original survey data. The residential neighborhood (urban, sub-urban, or rural) of the new household formed either due to marriage or divorce was obtained using probability rates observed in the original survey data. Also, we assumed that the children as well as all household resources including income, and household vehicles are split equally between the spouses in the event of a divorce. In cases where there is only one child and/or one vehicle in the household, we assigned the child/vehicle to one of the spouses randomly. In their recent survey of young adults in the U.S., the Pew Research Centre (PRC) found that nearly 40% of the adults aged 18-34 stay at home (PRC, 2012). We used this data to apply a move-out rate of 0.60 to all individuals when they reach an age of 18 years. Assuming that the young adult will move-out into either independent households or households with young roommates, we determined the characteristics of the household that the young adult moves into using probability rates calculated using proportion of households without older people (>=40 years).

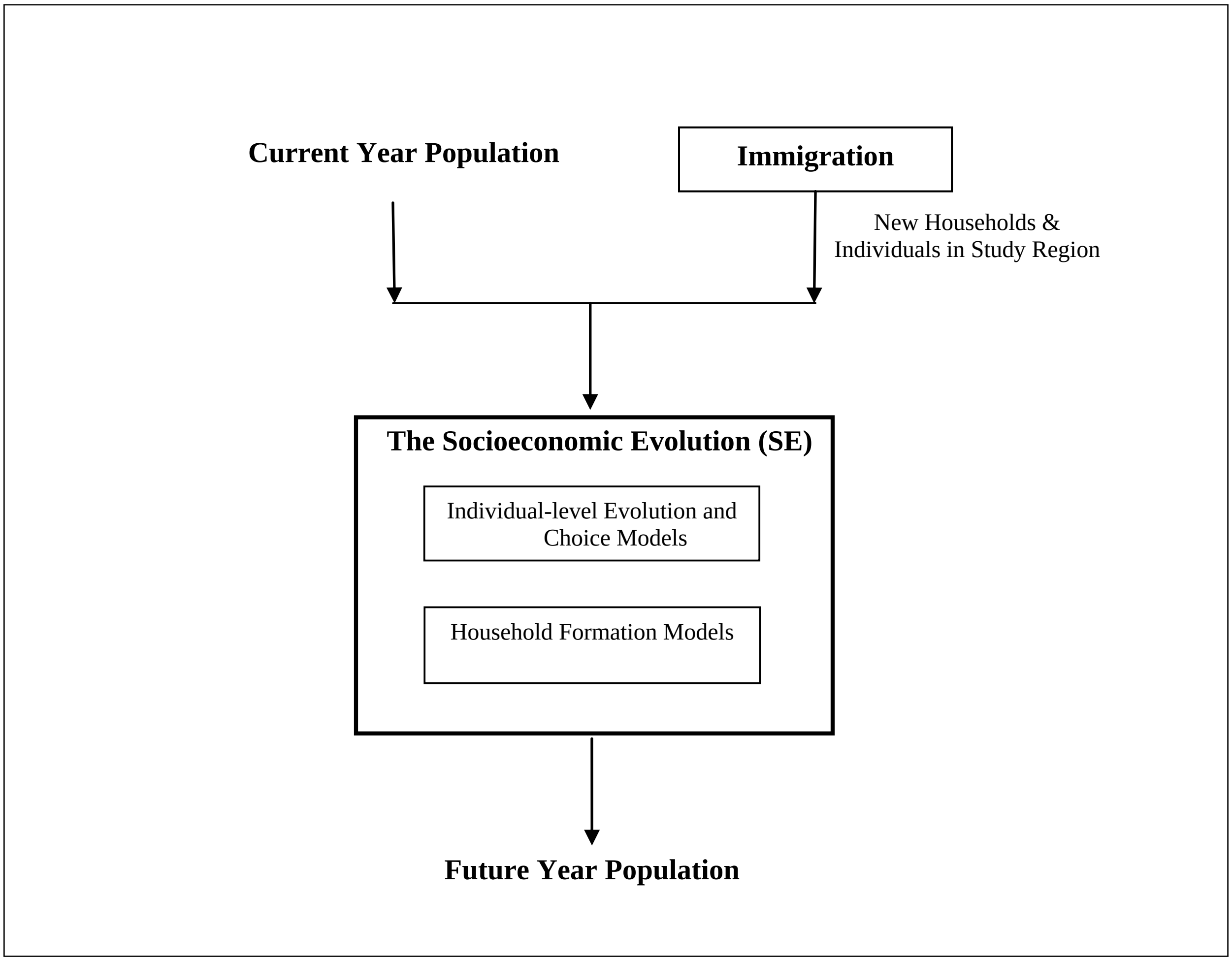

Figure 4.1 Overview of Population Evolution Framework

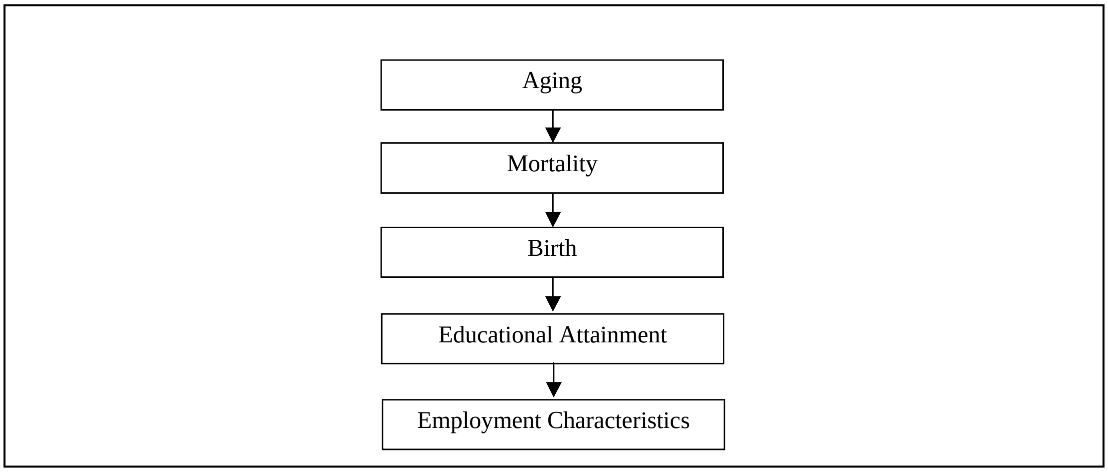

Figure 4.2a Individual-Level Evolution and Choice Models

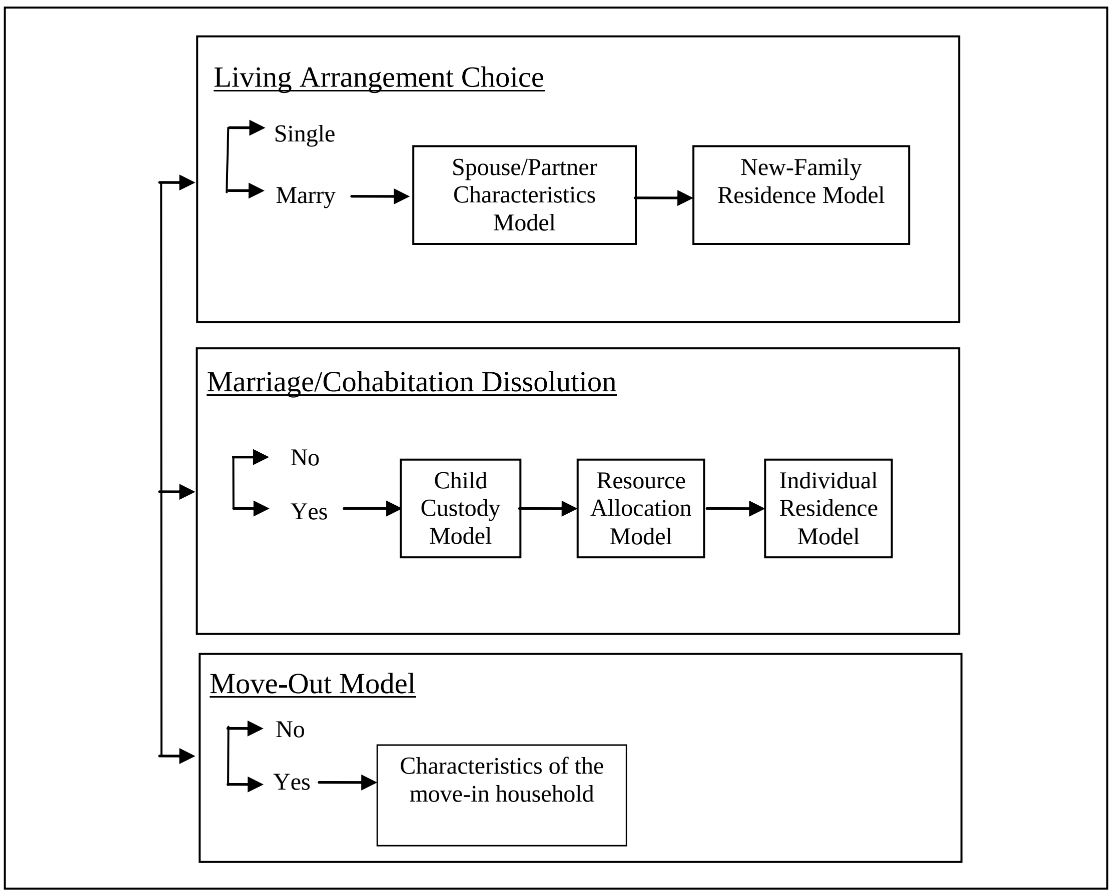

Figure 4.2b Household Formation Models

Figure 4.2 Detailed Overview of the Socioeconomic Evolution Framework

4.2.2 Vehicle Fleet Prediction

For predicting vehicle fleets, we start-off with the vehicle holdings observed in the data for the base year (*i.e.*, 2008). For each vehicle in the base year, we determine whether the household decides to keep, scrap, or replace the vehicle starting with the oldest vehicle in the vehicle fleet. If there is a scrap decision, the corresponding vehicle is removed from the fleet and the existing vehicle fleet characteristics are updated. Similarly, if there is/are replacement decision(s), then the corresponding vehicle(s) from the vehicle fleet are removed and the vehicle selection module is invoked to determine the characteristics of the new vehicle(s) that replaces (replace) the existing vehicle(s). After determining the transaction decisions of the existing vehicles, the household decision to purchase a new vehicle is simulated. If there is an "add vehicle" decision, then the vehicle selection module is invoked to determine the characteristics of the new vehicle.

For any new household created during the evolution, the synthetic choice occasions for each household are constructed based on the number of adults in the household. Then, we apply the vehicle selection module to determine the vehicle type (body type, vintage, and fuel type of the vehicle) and the associated annual mileage at each of the synthetic choice occasions, updating the vehicle fleet characteristics after each synthetic choice occasion. This process generates the vehicle fleet characteristics of all new households. This evolution procedure is continued year-by-year until the forecast year is reached. During this process, demographic variables are also appropriately evolved using the framework described in the above section. For the analysis of policy scenarios, the random seeds used in the microsimulation process for each household and for each choice decision occasion over the course of the forecasting period are held fixed at the base case values to ensure that any changes in the vehicle fleet characteristics and associated mileages are attributable to the policy under consideration.

4.2.3 Fuel Consumption and Greenhouse Gas Emissions Calculations

Our simulator predicts annual vehicle usage along with the vehicle type for each vehicle in a household. This enables the estimation of the total annual fuel consumption by dividing the annual mileage by a fuel economy (*i.e.*, mileage in miles per gallon) estimate based on the vehicle type. The average fuel economy value across all makes/models within each vehicle type (as defined by body type, vintage type, and fuel type) is used for the economy estimate for that vehicle type. For all vehicle types until model year 2012, we used the fuel economy data provided by the U.S. Department of Energy (available for download at http://www.fueleconomy.gov/feg/download.shtml). For all model years beyond 2012, we assume that new model years will come with 3% annual fuel economy increase. For example, if a new Gasoline car provides 35 mpg in 2012, we assume that a new Gasoline car in 2013 provides a mileage = 35*1.03 = 36.05 mpg. In addition, we also account for the proposed corporate average fuel economy (CAFE) standards for all light duty vehicles of model years 2017 to 2025 in our forecasts. Specifically, the new standards require all passenger cars (including sub-compacts to large sedans and station wagons, crossover utility vehicles, SUVs, and minivans) to have a minimum fuel economy of 37.8 miles per gallon (mpg) in model year 2012 and 56.0 mpg in model year 2025 and all light trucks to have a minimum fuel economy of 34.1 mpg in model year 2012 and 49.6 mpg in model year 2025 (NHTSA, 2012). So, for all new model years starting 2017 are set to meet these new CAFE standards in the case that their mileage computed using a 3% annual increase in fuel economy as described earlier comes out to be lower than the CAFE standard.

Also, we make a few reasonable assumptions in terms of when vehicles of different body types that use alternative fuels to gasoline will become available over time. For instance, fully electric large vans are not manufactured currently. However, we assume that these vehicles will become available starting 2015. Also, all the costs associated with vehicle usage including vehicle purchase price and vehicle maintenance cost are expected to increase by 3% every year. For hybrid-electric vehicles and plug-in hybrid-electric vehicles, we assume that the liquid fuel (which produces GHG emissions) used in the vehicles is gasoline (and not any other fuel such as diesel, or flex fuel). For example, if the economy value of a hybrid-electric vehicle is estimated to be 90 miles per gallon, under our assumption we treat the vehicle as providing 90 miles of travel per gallon of gasoline. The fuel economy value estimates of the compressed natural gas (CNG) and fully electric vehicles represent the miles per gallon of gasoline equivalent (MPGe) values. For CNG vehicles, a Gasoline Gallon Equivalent (GGE) factor of 0.51 cubic feet (at 3600 psi, which is the pressure in most CNG cylinders) is used to convert the gallons of gasoline to equivalent volume of CNG with the same energy content (US Department of Energy, 2008). All the fully electric vehicles emit zero mobile-source GHG emissions and thus are not considered in the GHG emissions calculation[3].

After this step, we obtain the total fuel consumption by gasoline, diesel, flex fuel, and CNG (since the liquid fuel in hybrid-electric and plug-in electric vehicles is assumed to be gasoline, they do not appear separately in the list of fuel types here). Then, the associated CO2 emissions are estimated using the following equation that EPA uses for all its emissions inventory calculations:

$$
\begin{aligned}
\mathrm{CO_2\ Emissions/Gallon} &= \mathrm{Carbon Content of\ Fuel} \\
&\quad \times \left(\frac{\mathrm{Molecular Weight of\ CO_2}}{\mathrm{Molecular Weight of\ Carbon}} \right) \times \mathrm{Oxidation\ Factor} \\
&= \mathrm{Carbon Content of\ Fuel} \ \mathrm{x} \left(\frac{44}{12} \right) \mathrm{x}\ 0.99
\end{aligned}
$$

The oxidation factor accounts for the fact that some percentage of carbon remains un-oxidized. The EPA suggests the use of an oxidation factor of 0.99. Also,the EPA uses 2,421 and 2,778 grams as the carbon content in gasoline and diesel vehicles (EPA, 2005). We assume that all flex fuel vehicles use an E85 blend that contains 85% ethanol and 15% gasoline. So, the carbon content of flex fuel is obtained as 2,421*0.15 = 363.15. The CO_2 emissions from CNG vehicles are computed using a carbon content value of 490 grams of carbon per cubic meter of CNG. This value is obtained from the Bio-energy Feedstock Information Network (BFIN) website (BFIN, 2012). All $non-CO_2$ GHG emissions including N_2O, CH_4, and HFC (hydrofluorocarbons) usually constitute 5% of

[3] In the current study, we consider only the tailpipe emissions that occur due to vehicle usage and not life-cycle emissions which include production and distribution emissions associated with the fuel.

total GHG emissions, so the total CO_2 emissions is multiplied by a factor of $\left(\frac{100}{95}\right)$ to obtain the total GHG emissions (EPA, 2005).

4.2.4 Policies Considered

State agencies across the U.S. are currently implementing many different GHG emissions control policies/strategies. These include HOV lane exemption for fuel efficient vehicles, parking incentives (such as allowing an individual to park an Alternate Fuel Vehicle (AFV) in areas designated for carpool operators), free parking on city streets for qualified AFVs and Hybrid Electric Vehicles (HEVs), reduced rental surcharges for electric vehicles, real property tax exemptions for electric vehicle charging systems, tax credits for costs of installing electric charging stations, alternate fuel equipment tax credit, rebate on HEV/AFV purchases, vehicle registration fee reduction/exemption, and reduced/exempted alternate vehicle fuel taxes. A detailed list of policies that are currently implemented in different states of the U.S. is available at: http://www.ncsl.org/default.aspx?TabId=19324 (National Conference of State Legislatures, 2011). In addition to the typical socio-demographic variables, the vehicle fleet simulator being developed in this study is sensitive to many of the policies just identified, making it an effective tool for comprehensive policy analysis.

The scenarios that will be evaluated in the current dissertation using the proposed vehicle fleet simulator may be grouped into three broad categories- (1) Incentive-based policies, (2) Future market conditions, and (3) Technological innovation based scenarios. The incentive-based policies for non-gasoline (CNG, hybrid-electric, plug-in hybrid, and fully electric) vehicles use include (a) *HOV lane access*, (b) *Free parking* in certain designated spots, (c) *$1,000 annual income tax credit*, (d) *50% reduced tolls*, and (e) *$1,000 reduced vehicle price.* Future market conditions include changes in the vehicle purchase price, maintenance cost, fuel cost, and fuel availability. Specifically, we focus on the fuel cost (gasoline cost doubles), fuel availability increased to 1 in 25 stations from 1 in 50 stations (this policy applies only to CNG and fully electric vehicles), and maintenance cost (25% reduction in annual maintenance cost) scenarios. Lastly, technological innovations include the development of (a) powerful vehicles that have lower acceleration times, (b) AFV/EV vehicles with increased driving range (this policy applies only to CNG and fully electric vehicles where all vehicles have more than 200 miles of driving range), and (c) vehicles with better fuel economy (the fuel economy of all vehicles is assumed increase by 6% annually). Specific selected combination scenarios are also considered in our analysis.

In the baseline scenario against which we will compare all the policy scenarios, we assume that new vehicle mileages go up by 1% every year, and vehicle purchase prices, fuel costs, and maintenance costs increase by 3% every year. We assume that the time-line when vehicles of different fuel types become available is the same in the baseline as well as policy scenarios. The impact of the policies on the total number of vehicles owned, vehicle fleet composition, consumption levels of different fuels, and the associated GHG emissions is then analyzed.

4.3 Results

4.3.1 Population Evolution

We compared the percentage increase in the population for the years 2020 and 2030 using the population evolution framework described in Section 4.1 with those provided by the California Department of Finance (DOF) for the year 2012 and with estimates by Pitkin and Myers, 2012. Table 4.1 presents the comparison results. It can be seen that our predictions are fairly close to the predictions by the DOF as well as Pitkin and Myers, 2012. Table 4.2 presents the population distribution by age for the years 2010, 2020, and 2030 obtained in this study with the Department of Finance estimates and those reported in Pitkin and Myers, 2012 which are the two most recent population estimates available for the State of California also based on the 2010 Census data. As can be noticed from the table, we are able to closely replicate the changing age structure (as predicted these other sources) in the population over years. Specifically, over the years, the percentage of young population (less than 18 years) is reducing to about 18% while the older age group

(65 and more) grows from 14.8% to in 2010 to 28.6% in 2030. The higher percentage of older age groups in our projections is probably due to the higher percentage of older (55-74 years) population in our base year population (*i.e.,* in the original survey sample) compared to the Census data. Overall, the results suggest that the population evolution framework seems to be performing reasonably well in terms of capturing the key trends in the population growth. However, as indicated earlier, there is considerable scope to incorporate more detailed models into the evolution framework to improve the accuracy of the forecasts.

Table 4.1 Population Growth

Year	Current Study	DOF, 2012	Pitkin and Myers, 2012
	Percentage Growth (from 2010)		
2010	--	--	--
2020	9.41	9.39	9.35
2030	18.53	19.46	19.90

Table 4.2 Population Distribution by Age

Age Group	2010			2020			2030		
	Current Study	DOF,[4] 2012	Pitkin and Myers, 2012	Current Study	DOF, 2012	Pitkin and Myers, 2012	Current Study	DOF, 2012	Pitkin and Myers, 2012
<10	11.3	13.5	13.5	10.4	12.1	12.1	10.3	12.5	11.4
10-17	9.7	14.5	11.5	8.2	13.1	10.5	7.7	11.8	9.5
18-24	9.5	7.4	10.6	7.8	6.9	9.6	7.6	6.5	9.1
25-34	10.1	14.3	14.4	14.5	15.0	15.2	11.2	13.7	14.0
35-44	12.5	13.9	13.8	10.0	13.1	13.4	14.5	14.0	14.4
45-54	16.2	14.1	14.0	11.5	12.4	12.4	9.8	11.8	12.2
55-64	15.9	10.8	10.9	14.5	12.3	12.0	10.4	10.9	10.9
65-74	9.2	6.2	6.1	13.2	8.9	8.8	12.4	10.2	9.9
75+	5.6	5.3	5.3	9.9	6.2	6.1	16.1	8.6	8.7
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

[4] The DOF, 2012 estimates are for the age groups 10-19 and 20-24 years.

4.3.2 Vehicle Fleet Forecasts

Table 4.3 presents the vehicle fleet characteristics in the years 2020 and 2030 under different scenarios. It can be seen from the table that under all scenarios, while the number of vehicles increases (a direct result of an increase in the driving population), the average number of vehicles per household decreases from the current level of about 2 vehicles per household to about 1.7 vehicles per household which is a nearly 15% drop in the vehicle ownership levels from the 2008 level. It appears that the average vehicle ownership drops by 0.1 every 10 years. In the scenario where the gas cost doubles, the average vehicle ownership drops further to about 1.6 which is about a 20% reduction from the 2008 level.

With regards to the vehicle fleet composition, there is a slight reduction in the percentage of compact cars and cars in general in the vehicle fleet. This is associated with a simultaneous increase in the percentage of larger vehicles, *i.e.* vans. This is a reflection of the tendency of households to own fewer but larger vehicles which can serve all members of the household. This is further substantiated by the slightly higher share of vans in the incentive based policy scenarios (*i.e.*, HOV lane access, tax credits, and reduced tolls). While we notice a drop in the pick-up trucks share from 2008 to 2020, it rises again from 2020 to 2030. This is a result of more fuel efficient options among pick-up trucks becoming available starting 2020 (we assume that AFV and hybrid electric options in pick-up trucks start becoming available starting 2020).

Next, we observe a significant reduction in the percentage of the vehicle fleet fueled by gasoline. Specifically, even in the baseline scenario (without any policies in place), the share of gasoline vehicles is found to reduce to about 45% in 2020 and to 40% in 2030. Hybrid electric, diesel, flex-fuel, and plug-in hybrid vehicles (in that order) are the fuel alternatives that gain significantly from these changing household preferences. While there is a more than 50% reduction in the gasoline share in the first 10 years, the corresponding reduction in the gasoline share is only 5% in the subsequent 10 years. This suggests that, some people in the population have a higher utility for gasoline vehicles irrespective of the technological innovations that might happen in the future, leading to a flattening demand profile of gasoline vehicles over time. With regards to fully electric vehicles, their share only increases to about 3.7% in the baseline scenario. People seem to prefer plug-in hybrid vehicles (nearly 12% market share in 2030 compared to 3.7% share of fully electric vehicles in the baseline scenario) over fully electric vehicles. While the incentive based strategies seem to slightly increase the share of fully electric vehicles to up to 6% (in the case of a tax credits scenario), the main boost to the fully electric vehicles seems to come from a technological innovation where all fully electric vehicles have a high driving range (*i.e.*, the estimated distance that one can drive with a fully shared battery) of more than 200 miles. In this hypothetical scenario, the share of fully electric vehicles increases to nearly 28%. This is also reflected in the 10% share of fully electric vehicles in 2030 under the better fuel economy scenario where we assumed that the driving range of fully electric vehicles increases by 6% annually. Also, the share of gasoline vehicles further reduces to 34% under these technological innovation based scenarios.

While incentive based scenarios alone are not able to trigger a greater shift towards fully electric vehicles, combination scenarios where incentives are coupled with a rise in gasoline prices can increase the share of fully electric vehicles to 11% in the scenario of providing tax credits to fully electric vehicles. With regard to the age composition of vehicle fleet, there is a clear shift towards newer vehicles in the future. To be specific, the share of new vehicles is found to saturate at about 30% in the future compared to less than 6% market share now. Better vehicle options with lower annual fuel costs seem to significantly influence the vehicle purchase decisions of households. Two interesting patterns can be noticed from the results corresponding to the technological innovation-based scenarios of increased driving range of electric and CNG vehicles and better fuel economy. Under the increased driving range scenario, the share of new vehicles is found to increase to up to 40%. This is reflection of people being very highly inclined towards acquiring new fully electric vehicles with high driving range. However, under the increased fuel economy scenario where fuel economy of all vehicles is assumed to increase by 6% annually, there is an almost equal distribution of the vehicle fleet across all vintage categories. So, there is not much to gain from the increased driving range of fully electric vehicles if all vehicles offer comparable fuel economies. Overall the results suggest promising trends that may lead to a greener and newer vehicle fleet in the future. However, the reality will be a combination scenario with several triggers arising due to incentives, market conditions, and/or technological innovations. Careful drafting of policies anticipating future market conditions can help realize the emissions goals within the next two decades.

Table 4.3 Vehicle Fleet Composition in 2020 and 2030

		Baseline		Incentive Based Scenarios									
				HOV Lane Access		Free Parking		Tax Credit		Reduced Tolls		Reduced Vehicle Price	
	2008	**2020**	**2030**	**2020**	**2030**	**2020**	**2030**	**2020**	**2030**	**2020**	**2030**	**2020**	**2030**
Total Number of vehicles	13016	15548	16697	15609	16908	15674	16840	15690	16918	15666	16909	15707	17064
Average Number of Vehicles per Household (miles)	1.98	1.84	1.71	1.84	1.73	1.85	1.72	1.85	1.73	1.85	1.73	1.85	1.74
Body Type	-												
Compact Car	22.24	23.65	21.56	23.72	20.59	24.46	20.71	23.92	20.70	23.91	21.08	24.55	21.33
Car	28.15	26.38	24.53	27.60	26.10	27.59	26.68	27.79	25.95	27.79	26.37	27.22	26.21
Small cross utility	5.57	8.67	6.97	7.34	5.50	6.87	5.24	6.71	5.14	7.19	5.64	6.90	5.35
Sports utility vehicle	20.02	19.93	19.88	20.43	19.62	20.68	19.89	20.52	19.78	20.46	19.16	20.53	18.69
Van	6.84	7.18	8.06	6.51	8.11	6.48	7.99	6.60	8.26	6.44	8.12	6.56	7.88
Pick-up truck	17.18	14.20	19.00	14.40	20.07	13.91	19.49	14.46	20.17	14.21	19.63	14.24	20.53
Fuel Type													
Gasoline	96.45	45.07	40.65	36.45	32.87	35.41	31.65	34.96	31.86	38.15	34.18	36.47	33.15
Flex Fuel	0.26	14.45	12.85	10.99	9.34	10.20	9.07	10.31	9.30	11.11	9.24	10.79	9.32
Plug-in Hybrid	0.02	9.48	12.01	14.01	17.88	14.51	18.23	14.82	18.63	12.88	16.53	13.89	17.38
CNG	0.07	0.49	0.43	0.67	0.75	0.68	0.94	0.80	0.75	0.70	0.74	0.81	0.64
Diesel	2.23	12.72	15.18	10.54	11.46	10.44	11.08	10.50	11.31	11.23	12.77	10.84	11.60
Hybrid Electric	0.93	15.08	15.22	22.79	22.16	24.01	23.42	23.88	22.04	21.77	21.08	22.66	22.13
Fully Electric	0.04	2.71	3.66	4.56	5.54	4.74	5.61	4.74	6.12	4.16	5.46	4.53	5.78
Vintage													
New	5.82	30.34	29.73	29.62	28.85	29.47	29.47	28.93	28.63	29.82	29.13	29.37	29.18
1-2 years	14.98	19.55	15.55	20.26	15.61	19.71	15.45	20.08	15.00	20.02	15.36	19.70	15.49
3-7 years	35.69	21.57	18.48	21.33	18.38	21.86	18.06	22.09	18.14	21.13	18.13	21.55	18.32
8-12 years	23.89	16.07	20.08	15.81	19.49	15.96	19.67	15.83	19.97	16.31	19.71	16.29	19.22
More than 12 years old	19.63	12.47	16.16	12.98	17.68	12.99	17.35	13.08	18.26	12.72	17.67	13.09	17.79

Table 4.3 Vehicle Fleet Composition in 2020 and 2030 (Continued)

		Future Market Conditions						**Technological Innovations**					
		Gas Cost Doubles		**Fuel Availability 1 in 25 Stations**		**Reduction in Maintenance Cost**		**Lower Acceleration Times**		**Increased Driving Range**		**Better Fuel Economy**	
	2008	**2020**	**2030**	**2020**	**2030**	**2020**	**2030**	**2020**	**2030**	**2020**	**2030**	**2020**	**2030**
Total Number of vehicles	13016	14832	15823	15411	16630	15689	16864	15720	16836	16091	17422	15805	17491
Average Number of Vehicles per Household (miles)	1.98	1.75	1.62	1.82	1.70	1.85	1.72	1.86	1.72	1.90	1.78	1.87	1.79
Body Type	-												
Compact Car	22.24	24.74	22.29	23.48	21.43	23.77	21.91	23.61	21.33	20.23	17.55	26.07	28.31
Car	28.15	26.15	25.47	26.14	24.38	26.22	24.21	26.06	24.10	23.11	21.91	26.03	26.73
Small cross utility veh.	5.57	9.20	7.63	8.72	7.18	8.52	7.23	8.35	7.03	8.69	7.15	8.50	7.10
Sports utility vehicle	20.02	19.34	18.42	20.41	19.86	20.55	19.81	20.77	20.41	19.95	17.98	19.42	15.54
Van	6.84	7.25	8.30	7.28	8.20	7.09	7.65	6.76	7.56	17.76	22.12	6.36	6.37
Pick-up truck	17.18	13.32	17.90	13.96	18.95	13.86	19.19	14.45	19.57	10.25	13.29	13.63	15.95
Fuel Type													
Gasoline	96.45	42.17	39.13	44.67	40.72	45.52	40.83	44.03	40.67	36.65	33.33	41.52	34.37
Flex Fuel	0.26	14.93	12.99	13.97	12.68	13.82	12.75	14.49	12.99	8.35	6.84	13.75	10.00
Plug-in Hybrid	0.02	9.57	12.07	9.42	11.85	9.18	12.29	9.32	11.97	5.23	6.98	10.81	14.65
CNG	0.07	0.80	0.86	0.50	0.61	0.49	0.59	0.45	0.45	3.36	3.74	0.35	0.33
Diesel	2.23	12.20	13.49	12.76	14.87	13.15	14.59	13.10	14.93	15.18	15.04	13.26	13.46
Hybrid Electric	0.93	14.49	14.78	15.15	15.27	14.93	15.60	15.55	15.62	8.17	7.00	16.08	17.20
Fully Electric	0.04	5.85	6.68	3.54	4.00	2.91	3.35	3.05	3.36	23.06	27.07	4.23	9.99
Vintage													
New	5.82	32.65	33.91	30.15	30.01	30.52	29.77	29.62	30.80	40.09	38.15	23.85	18.56
1-2 years	14.98	19.07	14.73	19.33	15.37	19.54	15.48	19.50	15.05	16.71	12.40	21.56	18.89
3-7 years	35.69	20.40	16.73	21.63	18.70	21.34	18.61	21.45	18.00	19.05	15.62	23.57	21.73
8-12 years	23.89	15.66	18.43	16.31	19.28	16.71	19.73	16.48	19.43	13.54	18.04	17.48	22.03
More than 12 years old	19.63	12.23	16.20	12.58	16.64	11.89	16.41	12.94	16.73	10.60	15.78	13.54	18.80

Table 4.3 Vehicle Fleet Composition in 2020 and 2030 (Continued)

		Combination Scenarios					
		Gas Cost Doubles + Tax Credit		Gas Cost Doubles + Fuel Availability 1 in 25 Stations		Gas Cost Doubles + Increased Driving Range	
	2008	2020	2030	2020	2030	2020	2030
Total Number of vehicles	13016	15023	15961	14692	15573	15536	16843
Average Number of Vehicles per Household (miles)	1.98	1.77	1.63	1.73	1.59	1.83	1.72
Body Type							
Compact Car	22.24	25.21	21.87	24.28	22.10	21.92	17.87
Car	28.15	28.67	27.39	26.80	25.88	24.85	23.84
Small cross utility vehicle	5.57	7.48	5.40	9.04	7.44	10.65	7.80
Sports utility vehicle	20.02	19.34	18.25	19.23	18.29	15.95	14.90
Van	6.84	6.68	8.42	7.13	8.74	17.68	23.94
Pick-up truck	17.18	12.63	18.66	13.52	17.56	8.95	11.65
Fuel Type							
Gasoline	96.45	31.49	29.13	41.48	38.62	29.06	27.22
Flex Fuel	0.26	10.62	8.54	14.61	12.95	6.67	5.25
Plug-in Hybrid	0.02	14.92	17.98	9.48	11.74	4.06	5.19
CNG	0.07	1.35	1.40	1.08	1.00	5.41	5.25
Diesel	2.23	9.31	9.95	12.35	13.62	11.88	11.98
Hybrid Electric	0.93	22.71	22.07	14.62	14.41	6.23	5.98
Fully Electric	0.04	9.60	10.93	6.38	7.67	36.69	39.13
Vintage							
New	5.82	31.72	31.49	32.88	33.00	39.04	36.13
1-2 years	14.98	18.87	14.37	18.70	14.54	17.26	12.68
3-7 years	35.69	21.19	17.60	20.55	17.69	19.50	15.12
8-12 years	23.89	15.57	19.22	15.48	19.01	13.99	18.51
More than 12 years old	19.63	12.65	17.33	12.39	15.75	10.21	17.56

4.3 Fuel usage and Emission Forecasts

Table 4.4 presents the vehicle mileage, fuel consumption, and emission predictions for 2020 and 2030. The average household mileage is predicted to increase by nearly 5% in the next 10 years before returning to near 2008 levels by 2030. Higher share of non-gasoline vehicles in 2030 with lower driving ranges might be pulling down household mileages and explaining the reductions in the average household mileage values beyond 2020. This is more clear from the results corresponding to the technological innovation-based scenario- an increased driving range where compared to the baseline scenario results in the average household mileage increasing by 4.7% and 4.3% in 2020 and 2030, respectively. Also, the result that household mileages are not reducing is interesting but not unexpected since fuel efficient options in the future are able to trigger a shift towards greener vehicles but are not necessarily taxing higher mileage accumulation. This is more evident from the results corresponding to the future market condition based scenario where the gas cost doubles. Under this scenario, there are 0.6% and 10% reductions in the average household mileage from 2008 levels in 2020 and 2030 since there is a higher cost associated with accumulating higher mileages.

In the baseline scenario, the average vehicle mileage is found to increase by 13% and 11% by 2020 and 2030, respectively. This is expected and consistent with the lower auto ownership levels in the future. In the baseline scenario, the total mileage is predicted to increase by about 35% and 42% by 2020 and 2030, respectively, from the 2008 level- consistent with the increase in the number of vehicles in the future. As expected, there is not much change in the future mileage levels in the incentive-based scenarios compared to the baseline scenario. However, under the market conditions based scenarios, the changes are significant and are along the expected directions. For instance, in the gas cost increase scenario, the total mileage increase by 2020 and 2030 is 5.4% and 5.8% less than the predictions in the baseline scenario.

In the technological innovation based scenario of increased driving mileage, the overall mileage in 2020 and 2030 is predicted to be 4.7% and 4.3% higher than the baseline scenario predictions. Gasoline fuel consumption is predicted to go down by 40%

and 47% by 2020 and 2030, respectively. This is associated with an increase in the consumption of other fuels including flex fuels, CNG, and diesel. Specifically, diesel consumption is predicted to increase significantly by nearly 370% by 2020. However, it is important to note that these predictions are based on our assumption that many clean diesel options become available in the future. This is consistent with the increasing share of diesel fueled vehicles discussed in the previous section. Although diesel vehicles offer higher mileage values, they have higher carbon content and thus produce more emissions. So, the overall impact of increase in the diesel market share on emissions will become clearer from the emissions predictions. The percentage reduction in gasoline fuel consumption is lower in all incentive-based scenarios compared to the baseline scenario. This is not entirely surprising since the liquid fuel in hybrid electric and plug-in hybrid vehicles, whose shares go up under these scenarios, is assumed to be gasoline.

Flex fuel and diesel consumption under all incentive-based scenarios is lower than the baseline scenario predictions consistent with the lower shares of flex fuel and diesel vehicles in these scenarios compared to the baseline scenario. On the contrary, CNG consumption is found to be higher in the incentive based scenarios compared to the baseline scenarios, again consistent with the higher share of CNG vehicles in these scenarios compared to baseline scenario. As expected, in the gas cost increase scenario, gasoline consumption is found to reduce by 11.7% and 13.4% more compared to the baseline scenario by 2020 and 2030, respectively. Technological innovations including increased driving range and better fuel economies induce much higher reductions in gasoline consumption than the gas cost increase scenario consistent with the lower share of gasoline vehicles in these scenarios compared to the gas cost increase scenario. Combination scenarios are much more effective and can result in up to 50% higher reductions in gasoline consumption compared to the baseline scenario. Expected trends in the consumption levels of other fuels are also observed.

Table 4.4 Vehicle Mileage, Fuel Consumption, and Emissions in 2020 and 2030

	2008	Baseline		Incentive Based Scenarios									
				HOV Lane Access		Free Parking		Tax Credit		Reduced Tolls		Reduced Vehicle Price	
		2020	2030	2020	2030	2020	2030	2020	2030	2020	2030	2020	2030
Average Annual Household Mileage (miles)	25.37	26.67	24.24	26.66	24.28	26.67	24.04	26.94	24.21	26.65	25.37	26.67	24.24
% change from 2008	--	5.12	-4.43	5.11	-4.28	5.11	-5.23	6.19	-4.55	5.05	-4.55	5.17	-3.44
% diff. from Baseline	--	--	--	-0.01	0.16	0.00	-0.84	1.02	-0.13	-0.07	-0.13	0.05	1.03
Average Annual Vehicle Mileage (in 1000 miles)	12.82	14.53	14.21	14.47	14.05	14.41	13.97	14.55	14.00	14.41	12.82	14.53	14.21
% change from 2008	--	13.36	10.82	12.91	9.61	12.44	8.95	13.48	9.23	12.43	9.29	12.26	9.55
% diff. from Baseline	--	--	--	-0.40	-1.09	-0.81	-1.68	0.11	-1.43	-0.82	-1.38	-0.96	-1.14
Total Mileage (miles)/10^6	166.85	225.92	237.19	225.91	237.56	225.91	235.19	228.23	236.88	225.77	236.88	226.03	239.63
% change from 2008	--	35.41	42.16	35.40	42.39	35.40	40.96	36.79	41.98	35.32	41.98	35.47	43.63
% diff. from Baseline	--	--	--	-0.01	0.16	0.00	-0.84	1.02	-0.13	-0.07	-0.13	0.05	1.03
Total Fuel Consumption: Gasoline (in gallons)/10^5	77.8	53.1	47.8	54.7	50.2	55	49.3	55.4	49.7	54.8	77.8	53.1	47.8
% change from 2008	--	-31.66	-38.59	-29.60	-35.41	-29.24	-36.62	-28.81	-36.15	-29.57	-36.71	-29.72	-35.40
% diff. from Baseline	--	--	--	3.01	5.17	3.54	3.20	4.17	3.96	3.05	3.05	2.83	5.20
Total Fuel Consumption: Flex Fuel (in gallons)/10^5	0.2	11.6	9.5	8.7	6.8	8.3	6.8	8.5	6.9	8.6	0.2	11.6	9.5
% change from 2008	--	5436.45	4433.10	4086.68	3167.29	3862.24	3172.48	3955.73	3204.80	4031.70	3276.44	3992.86	3218.17
% diff. from Baseline	--	--	--	-24.38	-27.92	-28.43	-27.81	-26.74	-27.10	-25.37	-25.52	-26.07	-26.80
Total Fuel Consumption: CNG (in gge)/10^5	0.1	0.4	0.3	0.6	0.6	0.5	0.8	0.7	0.6	0.6	0.1	0.4	0.3
% change from 2008	--	578.31	455.62	909.49	974.46	849.16	1338.26	1031.69	878.47	865.49	955.14	1056.33	756.31
% diff. from Baseline	--	--	--	48.82	93.38	39.93	158.86	66.84	76.10	42.34	89.90	70.47	54.12
Total Fuel Consumption: Diesel (in gallons)/10^5	1.9	10.2	11	8.5	8.6	8.2	8.1	8.8	8.5	9.2	1.9	10.2	11
% change from 2008	--	434.52	475.68	344.82	350.67	331.90	326.56	360.06	343.10	379.64	387.10	355.88	357.90
% diff. from Baseline	--	--	--	-16.78	-21.72	-19.20	-25.90	-13.93	-23.03	-10.27	-15.39	-14.71	-20.46
Total Emissions (grams)/10^9	73.99	60.44	55.76	59.16	54.51	59.07	53.17	59.91	53.71	60.04	73.99	60.44	55.76
% change from 2008	--	-18.31	-24.64	-20.05	-26.33	-20.16	-28.14	-19.04	-27.42	-18.85	-26.32	-19.84	-26.06
% diff. from Baseline	--	--	--	-2.13	-2.24	-2.27	-4.65	-0.89	-3.68	-0.66	-2.23	-1.87	-1.88

Table 4.4 Vehicle Mileage, Fuel Consumption, and Emissions in 2020 and 2030 (Continued)

	2008	Future Market Conditions						Technological Innovations					
		Gas Cost Doubles		Fuel Availability 1 in 25 Stations		Reduction in Maintenance Cost		Lower Acceleration Times		Increased Driving Range		Better Fuel Economy	
		2020	2030	2020	2030	2020	2030	2020	2030	2020	2030	2020	2030
Average Annual Household Mileage (in 1000 miles)	25.37	25.22	22.84	26.23	24.01	26.73	24.37	26.51	24.39	27.93	25.29	26.71	24.79
% change from 2008	--	-0.59	-9.95	3.40	-5.34	5.39	-3.92	4.50	-3.87	10.08	-0.30	5.30	-2.29
% diff. from Baseline	--	-5.43	-5.77	-1.64	-0.95	0.26	0.53	-0.59	0.58	4.72	4.32	0.17	2.24
Average Annual Vehicle Mileage (in 1000 miles)	12.82	14.41	14.12	14.42	14.13	14.44	14.14	14.29	14.17	14.70	14.20	14.32	13.86
% change from 2008	--	12.38	10.19	12.49	10.21	12.62	10.31	11.45	10.54	14.70	10.80	11.70	8.15
% diff. from Baseline	--	-0.86	-0.57	-0.76	-0.55	-0.65	-0.46	-1.68	-0.25	1.19	-0.02	-1.46	-2.40
Total Mileage (miles)/10^6	166.85	213.66	223.49	222.23	234.93	226.50	238.45	224.58	238.57	236.59	247.43	226.31	242.49
% change from 2008	--	28.06	33.95	33.19	40.80	35.75	42.92	34.60	42.99	41.80	48.30	35.64	45.34
% diff. from Baseline	--	-5.43	-5.77	-1.64	-0.95	0.26	0.53	-0.59	0.58	4.72	4.32	0.17	2.24
Total Fuel Consumption: Gasoline (in gallons)/10^5	77.80	46.90	41.40	52.20	47.10	53.20	48.20	52.80	47.60	37.20	31.00	32.40	18.30
% change from 2008	--	-39.65	-46.81	-32.84	-39.41	-31.55	-38.01	-32.13	-38.83	-52.14	-60.14	-58.31	-76.46
% diff. from Baseline	--	-11.70	-13.39	-1.73	-1.33	0.16	0.93	-0.69	-0.40	-29.97	-35.10	-39.00	-61.67
Total Fuel Consumption: Flex Fuel (in gallons)/10^5	0.20	10.80	9.00	10.50	9.20	10.70	9.40	11.40	9.60	6.80	5.30	6.80	3.00
% change from 2008	--	5070.05	4196.27	4916.13	4284.70	5004.73	4394.63	5357.94	4518.80	3143.73	2425.32	3139.07	1335.31
% diff. from Baseline	--	-6.62	-5.22	-9.40	-3.27	-7.80	-0.85	-1.42	1.89	-41.41	-44.29	-41.50	-68.34
Total Fuel Consumption: CNG (in gge)/10^5	0.10	0.60	0.60	0.40	0.50	0.30	0.50	4.00	0.30	3.00	3.20	0.20	0.10
% change from 2008	--	884.40	909.22	635.37	691.84	492.96	719.71	523.42	466.76	5113.54	5442.60	236.87	57.12
% diff. from Baseline	--	45.12	81.64	8.41	42.51	-12.58	47.53	-8.09	2.01	668.60	897.55	-50.34	-71.72
Total Fuel Consumption: Diesel (in gallons)/10^5	1.90	8.90	9.30	10.30	10.80	10.60	10.90	10.40	11.20	10.70	9.30	6.60	4.10
% change from 2008	--	368.29	386.90	437.20	464.27	457.29	473.14	446.92	487.02	459.25	386.77	243.51	113.42
% diff. from Baseline	--	-12.39	-15.42	0.50	-1.98	4.26	-0.44	2.32	1.97	4.63	-15.45	-35.73	-62.93
Total Emissions (grams)/10^9	73.99	53.33	48.13	59.56	55.00	60.90	56.11	60.37	55.87	46.05	38.48	37.13	21.06
% change from 2008	--	-27.93	-34.96	-19.51	-25.68	-17.70	-24.17	-18.41	-24.49	-37.77	-48.00	-49.81	-71.53
% diff. from Baseline	--	-11.77	-13.69	-1.47	-1.38	0.74	0.63	-0.12	0.20	-23.82	-31.00	-38.56	-62.23

Table 4.4 Vehicle Mileage, Fuel Consumption, and Emissions in 2020 and 2030 (Continued)

	2008	Combination Scenarios					
		Gas Cost Doubles + Tax Credit		Gas Cost Doubles + Fuel Availability 1 in 25 Stations		Gas Cost Doubles + Increased Driving Range	
		2020	2030	2020	2030	2020	2030
Average Annual Household Mileage (miles)	25.37	25.46	23.03	25.10	22.67	26.78	24.39
% change from 2008	--	0.37	-9.21	-1.05	-10.62	5.55	-3.85
% diff. from Baseline	--	-4.52	-5.00	-5.87	-6.48	0.41	0.61
Average Annual Vehicle Mileage (miles)	12.82	14.36	14.12	14.47	14.24	14.60	14.17
% change from 2008	--	12.02	10.13	12.92	11.12	13.91	10.53
% diff. from Baseline	--	-1.18	-0.62	-0.39	0.27	0.48	-0.26
Total Mileage (miles)/10^6	166.85	215.71	225.32	212.66	221.81	226.84	238.63
% change from 2008	--	29.29	35.05	27.46	32.94	35.96	43.02
% diff. from Baseline	--	-4.52	-5.00	-5.87	-6.48	0.41	0.61
Total Fuel Consumption: Gasoline (in gallons)/10^6	77.8	47.70	43.30	45.60	40.60	27.00	24.00
% change from 2008	--	-38.63	-44.27	-41.35	-47.81	-65.30	-69.10
% diff. from Baseline	--	-10.20	-9.26	-14.19	-15.01	-49.23	-49.68
Total Fuel Consumption: Flex Fuel (in gallons)/10^6	0.20	8.00	6.00	10.90	8.80	5.30	3.60
% change from 2008	--	3738.69	2771.63	5125.21	4128.71	2430.45	1623.77
% diff. from Baseline	--	-30.67	-36.65	-5.62	-6.71	-54.29	-61.97
Total Fuel Consumption: CNG (in gge)/10^6	0.10	1.00	1.00	0.90	0.70	4.50	4.20
% change from 2008	--	1673.22	1660.42	1443.35	1182.40	7692.70	7124.25
% diff. from Baseline	--	161.42	216.84	127.53	130.81	1048.83	1200.21
Total Fuel Consumption: Diesel (in gallons)/10^6	1.90	7.00	6.80	9.00	9.20	7.70	6.90
% change from 2008	--	266.26	257.97	372.78	381.13	305.77	259.34
% diff. from Baseline	--	-31.48	-37.82	-11.55	-16.42	-24.09	-37.58
Total Emissions (grams)/10^6	73.99	50.96	46.33	52.23	47.28	33.48	29.45
% change from 2008	--	-31.13	-37.39	-29.41	-36.10	-54.75	-60.20
% diff. from Baseline	--	-15.69	-16.91	-13.59	-15.21	-44.61	-47.19

In the baseline scenario, total emissions are found to reduce significantly in the future by nearly 18.3% and 24.6% by 2020 and 2030, respectively. The percentage reduction is not exactly proportional to the reduction in the gasoline vehicle market share. This is probably due to the increase in the number of diesel vehicles which produce more emissions (although more fuel efficient) in the future. This is more evident from the results corresponding to the incentive based scenarios where a reduction in the diesel market share is associated with a higher reduction of overall emissions. Incentive based scenarios such as free parking and tax credits are able to bring down the emissions by nearly 4.6 and 3.7% more than the baseline scenario. So, incentives might not be able to reduce vehicle mileages effectively but can bring down emissions by shifting people towards AFVs. Doubling of gas costs bring down the missions by nearly 11.8% and 13.7% more than the baseline scenario. Technological innovation based scenarios- increased driving range and better fuel economy are much more effective in reducing overall emissions. For instance, better fuel economy for all vehicles can reduce the overall emissions by 71.5% from the 2008 levels and by 62% more than the baseline scenario. However, it is important to note that this is purely hypothetical scenario where fuel economy of all vehicles is assumed to increase by 6% annually.

4.4 Conclusions

This study contributes in important methodological ways to the science of travel modeling. Specifically, it implements a comprehensive vehicle fleet evolution framework to model the entire vehicle fleet held by households by vehicle type and usage. This is a substantial improvement over earlier studies that have focused on vehicle type modeling for a single vehicle in the household and used rather aggregate characterizations of vehicle body type (such as cars and non-cars). Another important contribution of the proposed research is that it models the purchase decisions of alternative fuel and electric vehicle types that are only recently beginning to “hit” the market. This is because the CEC data used in estimating the model components of the vehicle fleet simulator collects prospective stated intention data about purchase of vehicles in the near future as well as stated preference data on the choice of a vehicle type from carefully designed experimental scenarios.

By using all the information from the CEC data, a microsimulation methodology to "evolve" vehicles over time is developed, which is sensitive to vehicle incentive-based policies, future market conditions, and technological innovation-based scenarios. The research contributes to the practice of travel modeling by applying the microsimulation platform to examine the effects of a suite of GHG control policies. The results from this study contribute to identifying effective strategies at the regional, state, and national levels to increase the penetration of alternative-fuel vehicles and fuel-efficient vehicles, reduce energy consumption, and reduce greenhouse gas emissions.

Overall, the results suggest that more than the incentive-based strategies, it is the gasoline fuel prices and future technological innovations that are going to drive people towards greener vehicles. Given that under some of the policy scenarios, the share of electric vehicles can increase substantially, it is important to undertake more comprehensive analysis of the additional burden that this shift would put on the electricity infrastructure of the region. Also, innovative strategies to meet the demand must be explored to facilitate this impending transition towards electric vehicles (Anderson *et al.*, 2009). However, one limitation of the analysis done in this study is that we focus only on the emissions associated with vehicle operation. For instance, although fully electric vehicles produce zero emissions during operation, there are emissions associated with the car manufacturing and the electricity generation. A more comprehensive well-to-wheel analysis is needed to understand the overall impact of these policies on the environment (Thomas, 2012). Also, given that the share of larger vehicles such as vans as well as new vehicles which will likely come with better safety equipment is predicted to increase in the future, it is important to explore how this would affect safety of road users. It is possible that aggressive driving behavior offsets the benefits offered by the modern safety equipment. For instance, one recent study Abay *et al.*, 2012 found that people offset the safety benefits of seat belt use by driving more aggressively.

CHAPTER 5: Spatial Vehicle Type Choice Model

The material in this chapter is drawn substantially from the following published paper:

Paleti, R., C.R. Bhat, R.M. Pendyala, and K.G. Goulias, (2012) The Modeling of Household Vehicle Type Choice Accommodating Spatial Dependence Effects. Forthcoming, *Transportation Research Record.*

This chapter presents a multinomial probit model formulation that incorporates spatial spillover effects arising from both observed and unobserved factors. Specifically, Section 5.1 presents a description of the data while Section 5.2 presents the modeling methodology employed in this study. Section 5.3 presents estimation results, and Section 5.4 offers an assessment of the model specification and inferred elasticity estimates. Section 5.5 summarizes the work and offers some concluding thoughts.

5.1 Data

The data set used in this study is derived from the California add-on component of the 2009 National Household Travel Survey (NHTS). The National Household Travel Survey (NHTS) is a national survey conducted by the United States Department of Transportation to measure the amount of personal travel that is undertaken by the nation's populace. Individual states and metropolitan areas are allowed to purchase and commission additional data collection within their jurisdictions if they desire larger samples for their own analysis and planning applications. Within the California add-on survey sample, the subsample from the Los Angeles city region was extracted for the analysis conducted in this research. As spatial interaction effects are likely to be more localized in nature, it was considered prudent to use a data set from a limited geographic region. The desire to limit the sample size (and thus avoid inflated t-statistics that might arise from the use of large samples) was another consideration in the selection of a subsample from a limited geographic region. Finally, the selection of this specific subsample made it possible to merge census tract level accessibility measures and land use data that have been compiled in connection with an ongoing parallel effort to develop a comprehensive activity-based microsimulation model system for the Southern California Association of Governments (Chen *et al.*, 2011). The accessibility measures are opportunity-based indicators which measure the number of activity opportunities by 12 different industry types as well as total roadway length of different roadway types that can be reached within 10 minutes using the auto mode from the home census tract during the morning peak period (6 AM to 9 AM).

The data set includes detailed individual and household level socio-economic and demographic data together with information about the vehicle fleet in each household. After extensive cleaning and filtering for missing data, a survey sample of 961 households was available for analysis. In order to limit the sample size and for reasons of computational tractability, a 25 percent random sample of 243 households residing in 200 census tracts was chosen for model estimation. For the model estimation exercise in this study, vehicle type choice was represented as a combination of two dimensions – body type and vintage. Two body types were considered, namely, car and non-car (encompassing sport utility vehicles, vans/minivans, and pick-up trucks). Two age categories were considered – less than or equal to five years old, and greater than five years old. Thus there are four vehicle type alternatives.

An examination of the descriptive characteristics of the sample of 243 households suggests that the data set is suitable for the model estimation effort undertaken in this study. It is found that 8.2 percent of households have no vehicle, another 34.5 percent have one vehicle, and 40 percent have two vehicles. Among the vehicles in the sample, 40 percent are old cars, 24 percent are new cars (less than or equal to five years old), another 24 percent are old non-cars, and 12 percent are new non-cars. Among other descriptive statistics, 82 percent of the households are of non-Hispanic origin, with 68 percent of individuals reporting their race as Caucasian. About 70 percent of households own the home in which they reside. With respect to the income distribution, it is found that one-fifth of the households report an annual income less than $20,000 and an equal proportion report incomes between $20,000 and $45,000. Just about 38 percent of the households report income greater than $75,000 per year. About 47 percent of the households report having one adult and another 46 percent report having two adults. Nearly 34 percent of households have zero workers, and 44 percent have one worker. About 17 percent of households report having one self-employed individual. There is one person with more than one job in 11 percent of the households. The employed individuals report a mean distance to work of 6.1 miles. Only one percent of the households report having a child 0-5 years of age, but 12 percent of households report having a child 6-10 years of age. About 12 percent of households report having a child 11-15 years of age (households not necessarily mutually exclusive). Just over one-third of households report having a senior adult who is 65 years of age or older. About 35 percent of households are immigrant households. The mean distance between census tracts, which is the distance measure used to capture spatial dependence effects due to proximity, is 11.1 miles with a standard deviation of 6.6 miles. Overall, the descriptive analysis of the sample indicated that the sample is suitable for model estimation.

5.2 Modeling Methodology

We use the same framework of synthetic choice occasions described in Chapter 3 in this study as well. However, we briefly describe this framework again here. The behavioral framework adopted in this study assumes that the observed vehicle fleet of a household is the result of a series of unobserved (to the analyst) repeated "synthetic" discrete choice occasions in which the household chooses not to purchase a vehicle or chooses a vehicle of a certain type (Eluru *et al.*, 2010). The number of synthetic choice occasions in such a "vertical" (over time) choice setting is linked to the number of driving age members in the household to exploit the fact that the number of vehicles owned by a household is virtually never greater than the number of driving age members (say N) plus two (in the data set used in the current analysis, 99.1% of households were covered by this condition). Thus, for each household, a set of N+2 synthetic choice occasions is created and an appropriate choice is assigned as the dependent variable. For estimation, there needs to be a procedure to assign a chosen alternative at each synthetic occasion. For this, the temporal sequence of vehicle purchases of the household, as reported in the survey, is used. For example, say a household owns an old sedan and a new sports utility vehicle (SUV), with the old sedan being purchased first. Then, the old sedan is the chosen alternative at the first choice occasion, and the new SUV is the chosen alternative in the second. The chosen alternative in the remaining two choice occasions is "no vehicle purchased". For the second choice occasion, information that the household already has an old sedan is used as an explanatory variable.[5] The procedure above mimics the dynamics of fleet ownership decisions, although there is no temporal component of the dynamics involved because synthetic choice occasions are considered; the observed information available is only that of vehicles held at a cross-sectional point in time with information on the sequence in which the currently held vehicles were purchased.[6]

5.2.1 Modeling Approach

Let the instantaneous utility U_{qti} of household q (q = 1, 2, …, Q) at synthetic choice occasion t (t = 1, 2, …, T_q) for vehicle type choice i (i = 1, 2, …, I; I = 5 in the empirical context of the current study, including the "no vehicle purchase" alternative) be a function of a (K×1)-column vector of exogenous attributes x_{qti} (including household demographics, types of vehicles "chosen" before the t^{th} choice occasion, and activity-travel environment characteristics). Let $T_q = N_q + 2,$ where T_q is the number of synthetic choice occasions for household q, and N_q is the number of driving age members in household q. Note that t does not have a chronological time interpretation. It is simply a device to accommodate multiple synthetic choice occasions and mimic the dynamics of fleet ownership decisions. That is, t=1 for a household A does not have any chronological time bearing to t=1 or t=2 for a neighboring household B. However, the choice occasions

[5] It is also possible to assign the old sedan to the first choice occasion, no vehicle in the second, no vehicle in the third, and the new SUV in the fourth occasion. However, both of these assignments give the same results, because the "dynamics" are based on what the household already owns in total, not what was chosen in the immediately previous choice occasion.

[6] This is as opposed to "true" vehicle fleet evolution models that analyze the dynamics of vehicle transaction decisions over time. The estimation of such models, while appealing from a behavioral standpoint, has been hampered by the paucity of longitudinal data on vehicle transactions. Moreover, many dynamic models have focused primarily on vehicle ownership (*i.e.*, transactions) with inadequate emphasis on the vehicle type, usage, and vintage considerations of the household fleet.

of different households may be considered to occur over a time period wherein households are interacting and exchanging utility signals. Thus, the spatial dependence across households is specified for each vehicle type *i* without any specific association to the choice occasion. That is, the utility U_{qti} for household *q* at choice occasion *t* for alternative *i* is related to the utility $U_{q't'i}$ for household *q'* and alternative *i* at each (and all) of the choice occasions t' (t'=1, 2 ,…, $T_{q'}$) of household *q'*. This is an important distinction from the traditional spatial dependency specifications for spatial panel discrete choice models, and leads to a specific form for the model in this study that has not appeared previously in the literature.

Thus, the utility U_{qti} incorporating a spatial lag structure is written as follows:

$$U_{qti} = \delta \sum_{q'} w_{qq'} \sum_{t'=1}^{T_{q'}} U_{q't'i} + \tilde{\alpha}_{qi} + \boldsymbol{\beta}'_q \mathbf{x}_{qti} + \tilde{\varepsilon}_{qti} \qquad (1)$$

where $w_{qq'}$ is a distance-based spatial weight corresponding to units *q* and q' (with $w_{qq} = 0$ and $\sum_{q'} w_{qq'} = 1$) for each (and all) *q*, δ $(0 < \delta < 1)$ is the spatial lag autoregressive parameter, $\tilde{\alpha}_{qi}$ is a normal random-effect term capturing a household-specific stationary preference effect for vehicle type *i*, and $\boldsymbol{\beta}_{\mathbf{q}}$ is a household-specific (K×1)-vector of coefficients assumed to be a realization from a multivariate normal distribution with mean vector **b** and covariance $\tilde{\boldsymbol{\Omega}} = \mathbf{LL}'$. Let $\boldsymbol{\beta}_{\mathbf{q}} = \mathbf{b} + \breve{\boldsymbol{\beta}}_{\mathbf{q}}$, where $\breve{\boldsymbol{\beta}}_q \sim MVN_K(0, \tilde{\boldsymbol{\Omega}})$ (MVN_K represents the multivariate normal distribution of dimension K). Also, write $\tilde{\alpha}_{qi} = \tilde{a}_i + \breve{\alpha}_{qi}$, and let the mean and variance-covariance matrix of the vertically stacked (I×1)-vector of random-effect terms $\tilde{\boldsymbol{\alpha}}_q \left[= \left(\tilde{\alpha}_{q1}, \tilde{\alpha}_{q2}, \ldots, \tilde{\alpha}_{qI} \right)' \right]$ be $\tilde{\mathbf{A}}$ and $\tilde{\boldsymbol{\Lambda}}$, respectively. $\tilde{\varepsilon}_{qti}$ in Equation (1) is a normal error term uncorrelated with $\tilde{\boldsymbol{\beta}}_q$ and all $\tilde{\alpha}_{qi}$ terms (*i* = 1, 2, …, *I*), and also uncorrelated across observation units *q* and synthetic choice occasions *t*. However, at each synthetic choice occasion *t* for household *q*, the $\tilde{\varepsilon}_{qti}$ terms may have a covariance (dependency) structure across vehicle types *i* due to choice-occasion unobserved factors that simultaneously increase or simultaneously decrease the utility of certain types of vehicles: $\tilde{\boldsymbol{\varepsilon}}_{qt}\left[=\left(\tilde{\varepsilon}_{qt1},\tilde{\varepsilon}_{qt2},\ldots,\tilde{\varepsilon}_{qtI}\right)'\right]\sim MVN_I(0,\tilde{\boldsymbol{\Psi}})$.

As usual, appropriate scale and level normalization must be imposed on $\tilde{\mathbf{A}}$, $\tilde{\boldsymbol{\Lambda}}$ and $\tilde{\boldsymbol{\Psi}}$ for identification purposes. Specifically, only utility differentials matter in discrete choice models. At the same time, whenever utility differentials are taken during estimation, they must all originate from the same underlying matrices $\tilde{\mathbf{A}}$, $\tilde{\boldsymbol{\Lambda}}$ and $\tilde{\boldsymbol{\Psi}}$. To achieve this, take the utility differentials with respect to the first alternative. Then, only the elements $\alpha_{qi1}=\tilde{\alpha}_{qi}-\tilde{\alpha}_{q1}\ (i\neq 1)$ and its covariance matrix $\boldsymbol{\Lambda}_1$, and the covariance matrix $\boldsymbol{\Psi}_1$ of $\xi_{qti1}=\tilde{\varepsilon}_{qti}-\tilde{\varepsilon}_{qt1}\ (i\neq 1)$, are estimable. So, a normalization $\tilde{\alpha}_{q1}=0\ \forall\ q$ is applied, implying that $\tilde{a}_1=0$. Also, develop $\boldsymbol{\Lambda}$ from $\boldsymbol{\Lambda}_{\mathbf{1}}$ by adding an additional row on top and an additional column to the left. All elements of this additional row and additional column are filled with values of zeros. Similarly, construct $\boldsymbol{\Psi}$ from $\boldsymbol{\Psi}_1$ by adding a row on top and a column to the left. This first row and the first column of the matrix $\tilde{\boldsymbol{\Psi}}$ are also filled with zero values. An additional normalization needs to be imposed on $\tilde{\boldsymbol{\Psi}}$ because the scale is also not identified. For this, normalize the element of $\tilde{\boldsymbol{\Psi}}$ in the second row and second column to the value of one. Note that all of these normalizations are needed for econometric identification purposes.

Next, define the following:

$\mathbf{U}_{qt}=(U_{qt1},U_{qt2},\ldots,U_{qtI})'$, $\tilde{\boldsymbol{\varepsilon}}_{qt}=(\tilde{\varepsilon}_{1t1},\tilde{\varepsilon}_{qt2},\ldots,\tilde{\varepsilon}_{qtI})'$ ($I\times 1$ vectors),

$\mathbf{U}_q=(\mathbf{U}'_{q1},\mathbf{U}'_{q2},\ldots\mathbf{U}'_{qT_q})'$, $\tilde{\boldsymbol{\varepsilon}}_q=(\tilde{\boldsymbol{\varepsilon}}'_{q1},\tilde{\boldsymbol{\varepsilon}}'_{q2},\ldots\tilde{\boldsymbol{\varepsilon}}'_{qT_q})'$ ($(T_q\times I)\times 1$ vectors),

$\mathbf{U}=(\mathbf{U}'_1,\mathbf{U}'_2,\ldots\mathbf{U}'_Q)'$, $\tilde{\boldsymbol{\varepsilon}}=(\tilde{\boldsymbol{\varepsilon}}'_1,\tilde{\boldsymbol{\varepsilon}}'_2,\ldots\tilde{\boldsymbol{\varepsilon}}'_Q)'$ ($RI\times 1$ vectors), $R=\sum_{q=1}^{Q}T_q$,

$\breve{\boldsymbol{\alpha}}_q=(\breve{\alpha}_{q1},\breve{\alpha}_{q2},\ldots,\breve{\alpha}_{qI})'$ (I×1 vector), $\breve{\boldsymbol{\alpha}}=\left[(1_{T_1}\otimes\breve{\alpha}_1)',(1_{T_2}\otimes\breve{\alpha}_2)',\ldots(1_{T_Q}\otimes\breve{\alpha}_Q)'\right]'$ ($RI\times 1$ vector), $\mathbf{x}_{qt}=(\mathbf{x}_{qt1},\mathbf{x}_{qt2},\ldots,\mathbf{x}_{qtI})'$ ($I\times K$ matrix), $\mathbf{x}_q=(\mathbf{x}'_{q1},\mathbf{x}'_{q2},\ldots,\mathbf{x}'_{qT_q})'$ ($(T_q\times I)\times K$

matrix), $\mathbf{x} = (\mathbf{x}'_1, \mathbf{x}'_2, \ldots, \mathbf{x}'_Q)'$ ($RI \times K$ matrix), and $\breve{\boldsymbol{\beta}} = \left(\breve{\boldsymbol{\beta}}'_1, \breve{\boldsymbol{\beta}}'_2, \ldots, \breve{\boldsymbol{\beta}}'_Q\right)'$ ($QK \times 1$ vector). Let $\mathbf{IDEN}_E$ be the identity matrix of size E, $\mathbf{1}_E$ be a column vector of size E with all of its elements taking the value of one, and $\mathbf{1}_{EE}$ be a square matrix of size E with all unit elements. Also, define the following matrix:

$$\tilde{\mathbf{x}} = \begin{bmatrix} \boldsymbol{x}_1 & 0 & 0 & 0. \cdots 0 \\ 0 & \boldsymbol{x}_2 & 0 & 0 \cdots 0 \\ 0 & 0 & \boldsymbol{x}_3 & 0 \cdots 0 \\ \vdots & \vdots & \vdots & \vdots \ldots \vdots \\ 0 & 0 & 0 & 0 \ldots \boldsymbol{x}_Q \end{bmatrix} (RI \times QK \ \ matrix), \tag{2}$$

Let $\tilde{R}_q = \sum_{j=1}^{q-1} T_j$, with the convention that $\tilde{R}_1 = 0$, and let $\breve{R}_q = \tilde{R}_q \times I$. Define a matrix $\mathbf{C}$ of size $RI \times RI$ that is filled with sub-matrices of size $(T_q \times I) \times (T_{q'} \times I)$ as follows: $[\mathbf{C}]_{\{(\breve{R}_q+1)-\breve{R}_{q+1}\},\{(\breve{R}_{q'}+1)-\breve{R}_{q'+1}\}} = w_{qq'} \otimes \mathbf{1}_{T_q T_{q'}} \otimes \mathbf{IDEN}_I$, where $[\mathbf{C}]_{\{(\breve{R}_q+1)-\breve{R}_{q+1}\},\{(\breve{R}_{q'}+1)-\breve{R}_{q'+1}\}}$ refers to the sub-matrix of $\mathbf{C}$ that corresponds to the $(\breve{R}_q + 1)^{th}$ through $\breve{R}_{q+1}{}^{th}$ rows and $(\breve{R}_{q'} + 1)^{th}$ through $\breve{R}_{q'+1}{}^{th}$ columns. Let $\mathbf{S} = [\mathbf{IDEN}_{RI} - \delta \mathbf{C}]^{-1}$ ($RI \times RI$ matrix) and $\breve{\mathbf{A}} = \left[(1_{T_1} \otimes \tilde{\mathbf{A}})', (1_{T_2} \otimes \tilde{\mathbf{A}})', \ldots (1_{T_Q} \otimes \tilde{\mathbf{A}})'\right]'$ ($RI \times 1$ matrix). Then, Equation (1) may be written in matrix notation as:

$$\mathbf{U} \ = \mathbf{S}\left[\breve{\mathbf{A}} + \mathbf{x}\mathbf{b} + \breve{\boldsymbol{\alpha}} + \tilde{\mathbf{x}}\breve{\boldsymbol{\beta}} + \tilde{\boldsymbol{\varepsilon}}\right]. \tag{3}$$

Let $[.]_e$ indicate the e^{th} element of the column vector [.], and let $d_{qti} = \breve{R}_q + (t-1)I + i.$ Equation (3) can be equivalently written as:

$$U_{qti} = \left[\mathbf{S}\{\breve{\mathbf{A}} + \mathbf{x}\mathbf{b}\}\right]_{d_{qti}} + \left[\mathbf{S}\{\breve{\boldsymbol{\alpha}} + \tilde{\mathbf{x}}\breve{\boldsymbol{\beta}} + \tilde{\boldsymbol{\varepsilon}}\}\right]_{d_{qti}}. \tag{4}$$

Define $V_{qti} = \left[\mathbf{S}\{\breve{\mathbf{A}} + \mathbf{x}\mathbf{b}\}\right]_{d_{qti}}$ and $\varepsilon_{qti} = \left[\mathbf{S}\{\breve{\boldsymbol{\alpha}} + \tilde{\mathbf{x}}\breve{\boldsymbol{\beta}} + \tilde{\boldsymbol{\varepsilon}}\}\right]_{d_{qti}}$. Household q chooses the vehicle type at synthetic choice occasion t that provides maximum utility. Let the "chosen" vehicle type (assigned as described previously) for household q at occasion t be m_{qt}. In the utility differential form, Equation (4) may be written as:

$$y_{qtim_{qt}} = U_{qti} - U_{qtm_{qt}} = H_{qtim_{qt}} + \xi_{qtim_{qt}};$$
$$H_{qtim_{qt}} = V_{qti} - V_{qtm_{qt}} \quad and \quad \xi_{qtim_{qt}} = \varepsilon_{qti} - \varepsilon_{qtm_{qt}};\ i \neq m_{qt} \tag{5}$$

Then stack the utility differentials $y_{qtim_{qt}}\ (= U_{qti} - U_{qtm_{qt}}, i \neq m_{qt})$ in the following order: $\mathbf{y}_{qt} = (y_{qt1m_{qt}}, y_{qt2m_{qt}}, \ldots, y_{qtIm_{qt}})'$, an $(I-1)\times 1$ vector; $\mathbf{y}_q = (\mathbf{y}'_{q1}, \mathbf{y}'_{q2}, \ldots, \mathbf{y}'_{qT_q})'$, an $[(I-1)\times T_q]\times 1$ vector; and $\mathbf{y} = (\mathbf{y}'_1, \mathbf{y}'_2, \ldots, \mathbf{y}'_Q)'$, an $[(I-1)\times R]\times 1$ vector. Correspondingly, let $\mathbf{H}_{qt} = (H_{qt1m_{qt}}, H_{qt2m_{qt}}, \ldots, H_{qtIm_{qt}})'$, an $(I-1)\times 1$ vector; $\mathbf{H}_q = (\mathbf{H}'_{q1}, \mathbf{H}'_{q2}, \ldots, \mathbf{H}'_{qT_q})'$, an $[(I-1)\times T_q]\times 1$ vector; and $\mathbf{H} = (\mathbf{H}'_1, \mathbf{H}'_2 \ldots, \mathbf{H}'_Q)'$, an $[(I-1)\times R]\times 1$ vector. It may be noted that $\mathbf{y}$ has a mean vector **H.**

To determine the covariance matrix of $\mathbf{y}$, several additional matrix definitions are needed. Define a matrix $\mathbf{\Lambda}$ of size $RI \times RI$ that is block-diagonal, with each block diagonal as follows: $[\mathbf{\Lambda}]_{\{(\bar{R}_q+1)-\bar{R}_{q+1}\},\{(\bar{R}_q+1)-\bar{R}_{q+1}\}} = \mathbf{1}_{T_qT_q} \otimes \tilde{\mathbf{\Lambda}}$ (q=1,2,….Q), $\mathbf{\Omega} = \tilde{\mathbf{x}}(\mathbf{I}_Q \otimes \tilde{\mathbf{\Omega}})\tilde{\mathbf{x}}'$ ($RI \times RI$ matrix), and $\mathbf{\Psi} = \mathbf{IDEN}_R \otimes \tilde{\mathbf{\Psi}}$ ($RI \times RI$ matrix). Let $\tilde{\mathbf{F}} = \mathbf{S}[\mathbf{\Lambda} + \mathbf{\Omega} + \mathbf{\Psi}]\mathbf{S}'$ and define **M** as an (q = 1, 2,…, Q): $[(I-1)\times R]\times[I \times R]$ block diagonal matrix, with each block diagonal having $(I-1)$ rows and I columns corresponding to the t^{th} synthetic choice occasion of household q. This $(I-1)\times I$ matrix for household q and choice occasion t corresponds to an $(I-1)$ identity matrix with an extra column of -1's added as the $m_{qt}{}^{\text{th}}$ column. Finally, the multivariate distribution of the utility differentials is obtained, $\mathbf{y} : \mathbf{y} \sim \boldsymbol{MVN}(\mathbf{H}, \mathbf{\Sigma})$, where $\mathbf{\Sigma} = \mathbf{M}\tilde{\mathbf{F}}\mathbf{M}'$. Next, let $\boldsymbol{\theta}$ be the collection of parameters to be estimated: $\boldsymbol{\theta} = [\mathbf{b}'; \text{Vech}(\tilde{\mathbf{\Omega}}); \tilde{\mathbf{A}}', \text{Vech}(\tilde{\mathbf{\Lambda}}), \text{Vech}(\tilde{\mathbf{\Psi}}), \delta]'$, where Vech($\tilde{\mathbf{\Omega}}$) represents the row vector of upper triangle elements of $\tilde{\mathbf{\Omega}}$. Then, the likelihood of the observed sample may be written succinctly as Prob[$\mathbf{y}^* < \mathbf{0}$].

$$L_{ML}(\boldsymbol{\theta}) = \text{Prob}[\mathbf{y}^* < 0] = F_{R\times(I-1)}(-\mathbf{H}, \mathbf{\Sigma}) \tag{6}$$

where $F_{R\times(I-1)}$ is the multivariate cumulative normal distribution of $R\times(I-1)$ dimensions. Despite advances in simulation techniques and computational power, the evaluation of such a high dimensional integral is literally infeasible using established estimation techniques.

5.2.2 Model Estimation Procedure

In view of the computational intractability of the likelihood function presented earlier, the current study uses Bhat's (2011) maximum approximate composite marginal likelihood (MACML) inference approach in estimation. The MACML approach combines a composite marginal likelihood (CML) estimation approach with an approximation method to evaluate the multivariate standard normal cumulative distribution (MVNCD) function. The CML approach works as follows. Instead of developing the likelihood of the entire sample, consider developing a surrogate likelihood function that is the product of the probability of easily computed marginal events. For instance, one may compound (multiply) pairwise probabilities of household *q* choosing the actual "chosen" vehicle type m_{qt} at occasion *t* and choosing the actual "chosen" vehicle type m_{qs} at occasion *s*, of household *q* choosing vehicle type m_{qt} at occasion *t and* household *q'* choosing vehicle type $m_{q's}$ at time *s*, and so on. The CML estimator is then the one that maximizes the compounded probability of all pairwise events. The CML function may be written as:

$$L_{CML}(\boldsymbol{\theta}) = \prod_{q=1}^{Q}\prod_{q'=q}^{Q}\prod_{t=1}^{T}\prod_{t'=t}^{T} \text{Prob}(C_{qt} = m_{qt}, C_{q't'} = m_{q't'}) \text{ with } q \neq q' \text{ when } t = t', \tag{7}$$

where C_{qt} is an index for the vehicle type chosen by household *q* at occasion *t*. Each of these pairwise probabilities is of $(I-1)\times 2$ dimensions, which may be computed easily using the MVNCD approximation method embedded in the MACML method. The pairwise marginal likelihood function of Equation (8) comprises $R(R-1)/2$ pairs of multivariate pairwise probability computations, which can itself become quite time consuming. Fortunately, in a spatial-temporal case where spatial dependency drops quickly with inter-observation distance, it should suffice to retain pairs within a certain threshold distance. This threshold value is estimated by testing different distance bands, starting from a small distance band and increasing the band. Then, the asymptotic variance matrix $V_{CML}(\hat{\boldsymbol{\theta}})$ is estimated for each distance band and the threshold distance value (say $\tilde{d}_{thresh}$) is chosen as the value beyond which there is either an increase or no additional decrease in the total variance across all parameters as given by $tr[V_{CML}(\hat{\boldsymbol{\theta}})]$ (*i.e.*, the trace of the matrix $[V_{CML}(\hat{\boldsymbol{\theta}})]$.

The CML estimator of $\boldsymbol{\theta}$ is consistent and asymptotically normal distributed with asymptotic mean $\boldsymbol{\theta}$ and covariance matrix given by the inverse of Godambe's (1960) sandwich information matrix (see Zhao and Joe, 2005):

$$V_{CML}(\hat{\boldsymbol{\theta}}) = [G(\boldsymbol{\theta})]^{-1} = [H(\boldsymbol{\theta})]^{-1} J(\boldsymbol{\theta})[H(\boldsymbol{\theta})]^{-1}, \text{ where} \tag{8}$$

$$H(\boldsymbol{\theta}) = E\left[-\frac{\partial^2 \log L_{CML}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right] \text{ and}$$

$$J(\boldsymbol{\theta}) = E\left[\left(\frac{\partial \log L_{CML}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial \log L_{CML}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right)\right].$$

The "bread" matrix $H(\boldsymbol{\theta})$ of Equation (8) can be estimated in a straightforward manner using the Hessian of the negative of the MACML likelihood function, evaluated at the MACML estimate $\hat{\boldsymbol{\theta}}$. On the other hand, the "vegetable" matrix $J(\boldsymbol{\theta})$ is not that straightforward to estimate. In the current study, the procedure suggested by Bhat (2011) is adopted.

In the estimations, the positive-definiteness of each of the $\tilde{\boldsymbol{\Omega}}, \tilde{\boldsymbol{\Lambda}}, \text{and}\, \tilde{\boldsymbol{\Psi}}$ matrices is guaranteed by writing the logarithm of the pairwise-likelihood in terms of the Cholesky-decomposed elements of these matrices, and maximizing with respect to these elements of the Cholesky factor. To ensure the constraint $0 < \delta < 1$, this term is parameterized as $\delta = 1/[1+\exp(\tilde{\delta})]$. Once estimated, the $\tilde{\delta}$ estimate can be translated back to obtain estimates of δ.

5.3 Estimation Results

This section presents results of the estimation of the multinomial probit model with spatial dependency effects on the California add-on data set of the 2009 National Household Travel Survey. Estimation results are presented in Table 5.1. A number of model specifications were estimated prior to arriving at the final model specification. A number of specifications involving mixing on all variables were tested; however, none of the mixing parameters including the random effects came out to be statistically significant in the final model specification. It is important to note that, even in the absence of mixing on variables, the model does not collapse to a cross-sectional spatial model. This is because the setup of the model is such that the utility associated with an alternative for one household at any given synthetic choice occasion is influenced by the utility associated with the same alternative across all synthetic choice occasions of all other households in the region, thus leading to a pseudo unbalanced panel setup due to unequal number of choice occasions across individuals. Another key finding is that there were no significant deviations in the error covariance matrix $\mathbf{\Psi}_1$ of $\xi_{qti1} = \tilde{\varepsilon}_{qti} - \tilde{\varepsilon}_{qt1}$ $(i \neq 1)$ from the corresponding matrix in an independent multinomial probit (MNP) model.[7] This finding implies that at any given synthetic choice occasion the choice occasion specific unobserved factors that influence the utility associated with different vehicle type alternatives are all independent and identically distributed.

Although the magnitudes of constants cannot be directly compared across different alternatives (because there are continuous variables in the utility formulations), the relative values may be loosely interpreted as providing an indication of the baseline preference for different vehicle types. It appears that new cars are the least preferred vehicle type while old cars are the most preferred; there is little difference in the baseline preference between old and new non-cars. It is found that households with higher levels of educational attainment are less inclined to acquire new non-cars. It is possible that these households are more environmentally conscious and savvy consumers shunning the expense and environmental consequences of driving new non-cars (sport utility vehicles, trucks, and vans). Those of Hispanic origin show a greater inclination to acquire old non-cars, while African Americans are less likely to acquire old cars and new non-cars.

[7] The covariance matrix of the error term differences in an independent MNP model has 1 and 0.5 as the diagonal and off-diagonal elements, respectively.

Table 5.1 Spatial Vehicle Type Choice Model Estimation Results

Variables	Old Car		New Car		Old Non-car		New Non-car	
	Coef	t-stat	Coef	t-stat	Coef	t-stat	Coef	t-stat
Constant	-0.0351	-0.19	-0.6960	-3.64	-0.5647	-1.92	-0.5432	-1.94
Demographics								
Highest Education Attainment in Household (Base is College degree or less)								
Bachelor Degree	--	--	--	--	--	--	-0.5185	-9.12
Post Graduate	--	--	--	--	--	--	-0.4172	-6.48
Hispanic Status (Base category is non-Hispanic)								
Hispanic Origin	--	--	--	--	0.4983	5.81	--	--
Race (base category is all other races)								
African American	-0.3183	-6.35	--	--	--	--	-0.1993	-4.22
Housing Tenure (Base category is rental home)								
Own	0.3690	9.48	0.6321	11.08	0.7265	11.71	1.0050	17.81
Household Income (Base category is all lower income levels)								
Greater than $75K	0.7863	12.91	0.6581	8.89	0.5848	8.29	0.8284	11.80
Number of adults	0.5195	15.57	0.3884	8.27	0.6625	14.02		
Number of full time workers	--	--	0.4139	8.24	--	--	0.4485	9.00
Number of people with more than one job	--	--	-0.3947	-4.12	--	--		
Mean Distance to work (in miles)	--	--	--	--	--	--	0.0115	9.97
Presence of children								
6 to 10 years	-0.2136	-4.48	--	--	--	--	--	--
11 to 15 years	--	--	--	--	0.1985	2.07	--	--
Presence of senior adults	--	--	--	--	-0.5534	-7.87	-0.5617	-8.18
Presence of individual with prolonged medical condition (> 5 yrs)	--	--	-0.3844	-9.21	--	--	--	--
Immigration Status (Base is non-immigrant household)								
Immigrant household	-0.3733	-6.39	-0.3733	-6.39	-0.1957	-3.62	-0.1957	-3.62
Existing vehicle fleet characteristics								
Number of old cars	-1.1195	-12.84	-0.9968	-18.36	-1.0363	-14.35	-0.8437	-12.52
Number of new cars	-2.0845	-32.36	-1.2579	-16.29	-2.3589	-9.24	-0.7159	-10.11
Number of old non-cars	-0.9627	-9.96	-0.7004	-11.51	-1.3573	-15.60	-0.4919	-10.52
Number of new non-cars	-1.5708	-34.02	-1.1944	-8.93	-1.9237	-30.98	-1.7878	-22.35
Accessibility Measures								
Primary arterial roads roadway length within 10 min. (in miles/10^4)	--	--	--	--	-1.7816	-12.15	--	--
Minor arterial roads roadway length within 10 min. (in miles/10^4)	--	--	--	--	-0.9369	-6.03	--	--
Collector roads roadway length within 10 min. (in miles/10^4)	--	--	--	--	0.5306	16.06	--	--
Total manufacturing employment that can be reached within 10 min. (/10^4)	--	--	-2.4896	-2.20	--	--	--	--
Total amount of arts employment that can be reached within 10 min. (/10^4)	-9.2334	-9.33	--	--	--	--	--	--
Spatial Interaction parameter (δ)	0.1872 (3.80)							

Households that own their home have a greater utility for all vehicle alternatives (in comparison to households that do not own their home), with a particular preference for new vehicles in comparison to older vehicles. As expected, households in the highest income category have a positive utility for all vehicle alternatives, with a higher preference for new non-cars and the lowest preference for old non-cars. However, the difference in magnitudes of coefficients across vehicle type alternatives is quite modest. As the number of adults increases, households are more likely to acquire old cars, new cars, or old non-cars – presumably these households have a greater need for multiple cars and show a greater disinclination to acquiring new non-cars in view of budget constraints. However, with the easing of budget constraints that invariably comes with greater number of workers in the household, it is found that households show a greater preference towards new cars or non-cars and shun older cars.

As the mean distance to work increases, households are more likely to acquire new non-cars presumably because people are looking for larger and more comfortable and reliable vehicles for the longer commute. The presence of children is associated with a smaller likelihood of acquiring older cars, and a greater likelihood of acquiring older non-cars. Presumably, such households prefer larger cars for the space, and newer cars for the reliability factor. Immigrant households have a lower utility across all vehicle type alternatives compared to non-immigrant households, but have a smaller negative coefficient on the non-car alternatives. Immigrant households may be located in more dense neighborhoods and may be more walk and transit oriented, thus contributing to the lower utility across all vehicle types. The coefficients on the non-car alternatives are less negative, probably because these households are of larger size motivating the acquisition of non-cars in preference to cars. Households with senior adults are less likely to acquire non-cars. There may be two explanations for this; first, seniors may have diminishing driving skills that make the driving and control of larger vehicles cumbersome, and second, seniors may be living in smaller households (empty nests) and so do not need larger vehicles (non-cars) any more.

As the temporal sequence in which vehicles were acquired in the household is known in the survey data set, information about the existing vehicle fleet was used as explanatory variables in the utility specification of vehicle type alternatives for all choice occasions after the first. This specification mimics the underlying dynamics in the purchase decisions with existing vehicles in the household influencing the vehicles that households acquire subsequently. As expected, parameter estimates for all vehicle types are negative suggesting that households increasingly choose to acquire "no vehicle" as they build up their vehicle fleet. The relative magnitudes of the coefficients can be used to draw inferences about how households tend to construct their fleets. It appears that households are somewhat variety seeking; for example, as the number of old cars increases, households are more prone to acquire new non-cars (least negative coefficient); as the number of new cars increases, households are more prone to acquire new non-cars (and shun older cars); as the number of old non-cars increases, households are more prone to acquire new non-cars or new cars; finally, as the number of new non-cars increases, households are more prone to acquire new cars. In all cases, the least negative coefficient is associated with a car type different from that representing the explanatory variable for the existing vehicle fleet.

With respect to accessibility measures, households with good access to primary and minor arterials have a lower preference for older non-cars suggesting that these households may be more auto-oriented (and hence located in census tracts with good roadway presence) and prefer to drive newer cars. Households in census tracts closer to manufacturing employment are less likely to acquire new cars, possibly because these census tracts are in lower income areas thus making new car purchases challenging. On the other hand, households in census tracts close to "arts" employment are less likely to acquire old cars; it is possible that these census tracts are in urban arts districts that are trendier and people who locate there are more likely to acquire newer cars.

5.4 Model Assessment

The model estimation effort yielded coefficient values that are largely reasonable and behaviorally intuitive. This section offers an assessment of the model from a number of different perspectives including the significance of the spatial dependency parameter, the goodness of fit of the model relative to a model that does not include spatial dependence, and differences in elasticity estimates between the multinomial probit model with spatial dependency and the independent multinomial probit model that ignores spatial dependency.

Among the different weight matrix specifications that were tested, the inverse distance based specification was found to offer the best fit. The spatial autoregressive parameter in the spatial lag formulation, δ, also turns out to be statistically significant with a value of 0.1872 and t-statistic of 3.80. This is evidence of the presence of spatial spillover effects arising either due to didactic interactions of individuals in proximally located households or due to residential self-selection effects that can lead to a clustering of households with similar vehicle type choice preferences.

Although the spatial parameter is statistically significant suggesting superior data fit in the spatial model compared to a corresponding non-spatial model, an alternative way to compare these nested models is through the adjusted composite likelihood ratio (ADCLRT) test (Bhat, 2011). The composite log-likelihood value for the non-spatial model is -138971.8 (52 parameters estimated) and that for the final spatial model is -138827.6 (53 parameters estimated). The ADCLRT test statistic of comparison between the two models is 7.84 which is greater than the critical chi-squared value of 3.84 associated with one degree of freedom, thus demonstrating presence of spatial interactions in vehicle type choice decisions. Ignoring such spatial interaction effects results in a model with poorer statistical goodness-of-fit.

Table 5.2 Aggregate-Level Elasticity Effects of the Spatial and Non-Spatial Models

Variables	No Vehicle		Old Car		New Car		Old Non-car		New Non-car	
	Non-spatial	Spatial	Non-spatial	Spatial	Non-spatial	Spatial	Non-spatial	Spatial	Non-spatial	Spatial
Demographics										
Highest Education Attainment in Household (Base is College degree or less)										
Bachelor Degree	3.43	2.90	5.20	4.05	8.53	7.34	5.58	4.97	-83.27	-69.29
Post Graduate	2.71	2.41	4.06	3.50	6.97	5.98	4.61	4.08	-66.52	-57.63
Hispanic Status (Base category is non-Hispanic)										
Hispanic Origin	-5.85	-14.51	-17.21	-39.66	-15.65	-36.84	84.69	197.79	-14.18	-31.90
Race (base category is all other races)										
African American	8.44	11.67	-45.50	-63.80	22.52	36.01	22.83	35.76	-25.34	-50.94
Housing Tenure (Base category is rental home)										
Own	-29.85	-35.61	1.94	7.49	45.89	60.13	50.88	59.74	101.92	91.74
Household Income (Base category is all lower income levels)										
Greater than $75K	-39.03	-43.62	56.80	70.00	45.09	40.81	11.23	11.81	88.77	98.06
Number of adults	-23.62	-68.07	34.20	80.83	15.90	20.32	69.54	234.10	-45.62	-88.07
Number of full time workers	-11.24	-45.83	-19.58	-65.85	66.15	240.55	-20.07	-66.73	79.67	312.38
Number of people with more than one job	4.49	7.65	10.02	18.13	-52.21	-92.91	9.83	17.08	13.23	24.81
Mean Distance to work (in miles)	-0.26	-0.59	-0.31	-0.78	-0.71	-1.46	-0.38	-0.77	6.13	13.38
Presence of children										
6 to 10 years	3.82	6.99	-26.48	-50.18	9.96	20.24	11.09	20.64	7.77	14.70
11 to 15 years	-1.96	-5.62	-5.77	-16.28	-5.36	-13.92	28.55	77.49	-4.87	-11.79
Presence of senior adults	9.07	8.62	22.45	21.88	23.76	25.47	-72.84	-68.95	-78.65	-79.93
Presence of individual with prolonged medical condition (more than 5 yrs)	4.55	7.07	10.14	16.62	-52.70	-85.69	9.88	15.62	13.23	23.29
Immigration Status (Base is non-immigrant household)										
Immigrant household	16.82	17.65	-28.88	-31.14	-33.84	-36.13	2.16	3.25	-5.46	-4.04
Existing vehicle fleet characteristics										
Number of old cars	45.81	81.34	-60.05	-98.99	-55.64	-97.71	-50.85	-97.31	-46.11	-93.50
Number of new cars	60.12	81.83	-91.74	-100.00	-57.61	-99.71	-93.96	-100.00	4.57	-85.41
Number of old non-cars	39.99	75.88	-50.98	-97.64	-32.43	-86.81	-80.39	-99.93	-8.86	-60.21
Number of new non-cars	61.30	83.14	-75.85	-99.98	-52.10	-99.26	-86.40	-100.00	-89.53	-99.97
Accessibility Measures										
Primary arterial roads roadway length within 10 min. (in miles)	1.81	4.34	4.89	13.70	4.56	12.41	-25.15	-64.82	4.38	11.60
Minor arterial roads roadway length within 10 min. (in miles)	1.18	2.90	3.26	8.89	2.90	7.82	-16.35	-42.05	2.70	7.15
Collector roads roadway length within 10 min. (in miles)	-4.51	-16.76	-10.89	-38.14	-10.52	-34.42	58.78	203.70	-9.49	-31.05
Total amount of manufacturing employment that can be reached within 10 min.	0.23	0.55	0.45	1.08	-2.44	-6.03	0.42	1.02	0.57	1.60
Total amount of arts employment that can be reached within 10 min.	0.65	1.84	-4.41	-12.69	1.83	4.95	1.51	4.92	1.44	3.91

A question that is often raised in the context of advanced choice models that incorporate additional (observed or unobserved) effects is the extent to which policy forecasts might actually differ due to ignoring such effects. Although the goodness-of-fit is significantly better and the spatial interactions parameter is significant, does that mean that policy forecasts would be different as a result of using one model versus the other (that ignores spatial dependence effects)? The parameter estimates in Table 5.1 do not directly provide the magnitude of the impact of variables on the probabilities of acquiring each vehicle type alternative. To shed light on this question, it is useful to compute aggregate-level elasticity effects of variables for the different model specifications. Specifically, the effects of variables on the expected share of each vehicle type alternative are examined in this study. This is achieved by computing the marginal probability of each household choosing a certain vehicle type in a single synthetic choice scenario and aggregating these probabilities across households and all choice occasions for each vehicle type alternative.

The following procedure is used to compute the shares of each vehicle type alternative. The utility function of vehicle type i for household q is as follows:

$$U_{qti} = \delta \sum_{q'} w_{qq'} \sum_{t''=1}^{T_{q'}} U_{q't'i} + \boldsymbol{b}'\boldsymbol{x}_{qti} + \tilde{\varepsilon}_{qti} \qquad (9)$$

where the notation is similar to that described in the methodology section of this chapter. Then, using other notations described previously, it is possible to write:

$$\mathbf{U} = \mathbf{S}[\mathbf{x}\mathbf{b} + \tilde{\boldsymbol{\varepsilon}}], \qquad (10)$$

The above $RI \times 1$-vector $\mathbf{U}$ is simulated 500 times using the estimated values of $\mathbf{b}$, and by randomly drawing 500 times from the appropriate normal distributions for $\tilde{\boldsymbol{\varepsilon}}$. Next, the chosen alternative is determined as the alternative with the highest utility for each of the 500 draws. Finally, the predicted share of each alternative across the 500 draws is taken as an estimate of the probability of each vehicle type alternative. The aggregate share (across all households and all synthetic choice occasions) of each vehicle type alternative is obtained by aggregating the synthetic choice occasion level probabilities of each vehicle type alternative across all households.

The elasticity computed is a measure of the percent change in the aggregate share of each vehicle type alternative due to a change in an exogenous variable. For dummy variables, the value of the variable is changed to one for the subsample of observations for which the variable takes a value of zero, and to zero for the subsample of observations for which the variable takes a value of one. The shifts in expected aggregate shares in the two subsamples are then added after reversing the sign of the shifts in the second subsample, yielding the effective percent change in the expected shares across all households in the sample due to a change in the dummy variable from 0 to 1. For continuous variables, the value of the variable is increased by 25 percent for each observation and the percent change in the expected shares is computed. For variables which only take integer values (such as number of full time workers), the value is increased by unity.

Elasticity estimates are computed for the non-spatial MNP and the spatial MNP model, and are presented in Table 5.2. The first entry in Table 5.2 indicates that, according to the MNP model with no spatial interaction, households with a highest education attainment of Bachelor's degree are 3.43% more likely to not acquire a vehicle at any given choice occasion compared to other households. Other elasticity effects can be interpreted similarly.

All of the elasticity effects are consistent with the parameter estimates in Table 5.1. Also, the elasticity effects of the spatial and non-spatial models are in the same direction (sign) for all variables. However, the elasticity estimates of the non-spatial MNP model and spatial MNP models are quite different *in magnitude*. In general, the elasticity effects of the spatial model are consistently higher in magnitude than those from the non-spatial model. For instance, the elasticity effect of the number of full time workers for the new non-car alternative is 312% where as the corresponding number according to the non-spatial MNP model is only 80%. Similarly, the spatial model implies that Hispanic households are 198% more likely to obtain an old non-car whereas the non-spatial model implies only 85% higher likelihood for Hispanic households. Although the magnitude of the spatial autoregressive parameter is relatively small, the spatial spillover effect is compounding the elasticity estimates due to the circular reinforcing mechanism whereby a change in the value of a variable for one household changes utilities of vehicle type alternatives for other nearby households, which in turn causes ripple effects in the utility values of the household for which the variable changed in the first place.

In summary, it can be seen that elasticity estimates differ substantially between a model with and a model without spatial dependency effects. These differences can have dramatic implications for policy forecasts which often rely on model parameter estimates to infer the magnitude of change in behavior in response to a change in input conditions.

5.5 Conclusions

The important role played by vehicle ownership, fleet composition, and utilization in the energy and environmental arena underscores the need to develop rigorous models capable of accurately forecasting these choice dimensions for households in any geographic region. Although there has been considerable research on and progress in the development of vehicle ownership and fleet composition and utilization models, much of the previous work has neglected to fully incorporate spatial interaction effects that can shape household vehicle choices. Households are likely to interact with and observe the choices made by other nearby households, and this interaction may influence their own choices. As the distance between households increases, the level of spatial dependence presumably drops. In addition to spatial spillover effects due to observed factors of proximally located households, there may also be spatial error correlations arising from unobserved factors such as the attitudes and lifestyle preferences of households that may bring about residential self-selection. In the presence of spatial spillover effects (due to observed or unobserved factors), parameter estimates of discrete choice models of vehicle ownership will be inconsistent leading to potentially erroneous policy forecasts.

This study presents a multinomial probit model of vehicle ownership by type (fleet composition) that explicitly incorporates spatial interaction effects due to observed and unobserved factors. The model is estimated on the Los Angeles region subsample of the California add-on data set of the 2009 National Household Travel Survey that includes a host of accessibility and land use variables critical to vehicle ownership modeling. Underlying the model is a behavioral framework that considers the household vehicle fleet as being constructed over a series of purchase choice occasions, thus providing the ability to endogenously determine the vehicle fleet size while simultaneously incorporating history dependency in the choice model. In other words, the vehicle type that is acquired at any choice occasion is dependent on the existing vehicle fleet in the household comprising vehicles that were acquired at earlier choice occasions. The model considers five choice alternatives for each occasion – two body types (car and non-car) by two vintage types (less than or equal to five years old, and greater than five years old) plus the choice of acquiring no vehicle at all. The maximum approximate composite marginal likelihood (MACML) estimation procedure is employed to overcome computational intractability associated with traditional simulation and Bayesian model estimation procedures.

Model estimation results show that a host of individual and household variables, not to mention accessibility and land use variables, significantly impact choice of acquiring different vehicle types. More importantly, in the context of this study, it is found that the spatial interaction parameter is statistically significant and the model that incorporates spatial spillover effects offers a superior statistical goodness-of-fit compared to a multinomial probit model that does not incorporate spatial dependency effects. It is found that a distance based spatial interaction function offers the best fit, with interaction between households dropping off as the distance between households increases. A comparison of elasticity estimates offered by the spatial effects choice model estimated in this study against those offered by a model with no spatial effects shows that elasticity estimates differ substantially when spatial effects are incorporated. The elasticity estimates from the spatial effects model are consistently higher in magnitude, suggesting that interaction effects amplify the extent to which households modify their behavior in response to changes in input conditions. Incorporating spatial effects in models of discrete choice behavior can result in substantially different policy forecasts, with clear implications in the transportation planning and policy arena. Future research efforts in this domain could further explore the use of alternative spatial interaction functions, examine whether the findings hold true in other geographical contexts and data sets, and attempt to separate unobserved and observed spatial interaction effects by jointly modeling residential location choice with vehicle ownership and fleet composition choice.

CHAPTER 6: Integrated Tour Based Model System

The material in this chapter is drawn substantially from the following technical paper:

Paleti, R., R.M. Pendyala, C.R. Bhat, K.C. Konduri, (2011) A joint tour-based model of tour complexity, passenger accompaniment, vehicle type choice, and tour length. Technical Paper, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin.

As mentioned earlier in Chapter 2, four choice dimensions related to daily activity-travel choices are considered this study. They are tour complexity, passenger accompaniment, vehicle type choice, and total tour length. This chapter provides the definition of each of these attributes considered in this research and describes the multi-dimensional modeling framework used for the analysis. Specifically, Section 6.1 explains each of the four attributes considered in this study in more detail. Section 6.2 provides methodological details, identification, and positive definiteness considerations associated with the integrated modeling system developed in this study. Section 6.3 describes the survey data and the final estimation sample characteristics used for the analysis. Section 6.4 discusses the empirical results and Section 6.5 finally concludes the chapter summarizing the key findings and contributions of the current study.

6.1 Multi-dimensional Modeling of Tour Attributes

For the model development exercise of this research, tour complexity is represented by the number of stops made on the tour. A tour itself is defined as a closed chain, with the beginning and ending of the tour being the home location. Only home-based tours are considered in this study because of the desire to model vehicle type choice and it is presumed that one would have a choice among vehicle types (in a multiple vehicle household) when a tour begins at home. Stop frequency could be represented as an ordered response variable (Bhat and Srinivasan, 2005); however, within the context of this study, stop frequency is represented as a binary choice variable between the choice of making a one-stop tour or a multiple stop tour. The former may be considered "simple" tours while the latter may be considered "complex" tours. This simplification was done because of the generally low frequency of multiple stop tours in travel survey data sets.

Passenger accompaniment is a variable of much interest because it captures multiple behavioral processes at play. Passenger accompaniment is representative of joint or solo activity engagement, and thus captures (at least in part) interactions among household members. There is increasing recognition of the importance of intra-household interactions in modeling daily activity-travel patterns due to the inevitable linkages and dependencies that exist (Zhang *et al.*, 2005). Children, for example, are often dependent on parents for meeting travel needs (Paleti *et al.*, 2011). Household members often undertake activities jointly, particularly in the context of maintenance and discretionary activities, and this jointness in activity engagement may have important implications for destination choice (tour length), vehicle type choice, and time of day choice. In this study, passenger accompaniment is represented as a ternary choice variable with possible choice options being a pure solo tour, a pure joint tour (with multiple vehicle occupants throughout the tour), and a partly joint tour (with a single occupant for a part of the tour, and multiple occupants for the other part of the tour).

Vehicle type choice is a variable of much importance and considerable interest from an energy consumption and environmental assessment perspective. However, there is a paucity of research that explicitly addresses vehicle type choice in the context of tour-based model systems. There is a vast body of literature devoted to modeling vehicle ownership. While early research focused heavily on modeling the count of vehicles (Mannering and Winston, 1985), more recent work has provided frameworks for modeling vehicle fleet composition of households with vehicle types defined by body type, make and model, fuel type, and vintage (Bhat and Sen, 2006). In addition, there have been numerous studies that have attempted to model vehicle holding durations, and the timing and nature of vehicle transactions including acquisition, disposal, and replacement of vehicles (Mohammadian and Miller, 2003; Yamamoto *et al.*, 1999). Thus, while there is a base of research that offers methods to model and forecast vehicle ownership by type, there is virtually no research that subsequently uses that information in the activity-travel microsimulation process. Vehicle allocation to drivers, and the choice of vehicle for individual tours, are not dimensions that are explicitly simulated, thus limiting the ability to exploit the detailed information output from activity based microsimulation models in estimating energy consumption and emissions inventories. For this reason, the current study includes vehicle type choice as one of the dimensions in the system. In this study, for simplicity, vehicle type is represented as a multinomial choice variable with the universe of options being car (auto), sports utility vehicle (SUV), van/minivan, and pick-up truck.

The final choice dimension that is captured in the model system of this research is destination choice. There is a rich body of evidence on destination choice behavior and spatial processes at play in how people perceive spatial opportunities and choose destinations, while considering the time-space and institutional constraints that govern such choices (Pendyala *et al.*, 2002; Bhat and Zhao, 2002). However, destination choice is a dimension that applies to the individual trip or stop level, and not the tour level, because tours may have multiple destinations associated with multiple stops. In this study, total tour length is used to capture distances traveled in reaching the one or more destinations on a tour. Total tour length is a continuous variable and is of much interest because it is representative of vehicle miles of travel (VMT), a travel model outcome that is used to quantify total travel and assess impacts on energy and emissions estimates (Shiftan and Suhrbier, 2002).

Thus, the model system in this study combines a binary choice variable with two multinomial choice variables and one continuous variable. One of the two multinomial choice variables (vehicle type choice) has varying choice sets across choice-makers, depending on the vehicle fleet of the household to which the traveler belongs. The modeling of such a mixture of dependent variable types in a single integrated model system is quite complex, and this dissertation presents a state-of-the-art methodological framework for doing so by exploiting some of the recent developments in choice modeling and estimation methods. The methodology involves the deployment of new estimation techniques that reduce the dimensionality of the problem, thus eliminating some of the concerns associated with computational burden and imprecision that might arise when adopting simulation-based estimation approaches in the context of large multi-dimensional problems (Bhat, 2011).

Overall the earlier literature on integrated modeling of activity-travel choices illustrates the level of interest in the choice dimensions considered in this research and the need for advances in multi-dimensional integrated choice modeling that would allow the profession to recognize the package or bundle nature of multiple choice processes. While the model system may appear to be a theoretical effort at exercising econometric complexity, the model specification, formulation, and estimation approach presented in this study offers the potential for dramatic breakthroughs in activity-based travel demand modeling.

6.2 Modeling Methodology

This section presents a detailed description of the modeling methodology developed for estimating a multi-dimensional choice model system involving a mixture of dependent variable types. The model formulation accounts for correlated unobserved factors affecting multiple choice dimensions, and allows the estimation of all model parameters in a single step akin to classic full-information maximum likelihood approaches, thus ensuring the use of all information in parameter estimation leading to gains in efficiency. The remainder of this section presents the formulation.

6.2.1 Model Framework

Let there be G nominal variables for an individual, and let g be the index for the nominal variables (g =1, 2, 3,…,G)[8]. In the empirical context of the current research, G=3 (the nominal variables are accompaniment type, tour type or complexity, and vehicle type). Also, let I_g be the number of alternatives corresponding to the g^{th} nominal variable ($I_g \geq 2$) and let i_g be the corresponding index (i_g = 1, 2, 3…... I_g). Note that I_g may vary across individuals, but index for individuals is suppressed at this time for ease of presentation.

[8] A nominal variable can be an unordered multinomial response variable or a binary response variable.

Also, it is possible that some nominal variables do not apply for some individuals, in which case G itself is a function of the individual q. However, the model is developed at the individual level, and so this notational nuance does not appear in the presentation here.

Consider the g^{th} nominal variable and assume that the individual under consideration chooses the alternative m_g. Also, assume the usual random utility structure for each alternative i_g.

$$U_{gi_g} = \boldsymbol{\beta}_g' \boldsymbol{x}_{gi_g} + \varepsilon_{gi_g}, \tag{1}$$

where $\boldsymbol{x}_{gi_g}$ is a (K_gx1)-column vector of exogenous attributes, $\boldsymbol{\beta}_g$ is a column vector of corresponding coefficients, and ε_{gi_g} is a normal error term. Let the variance-covariance matrix of the vertically stacked vector of errors $\varepsilon_g \left[= \left(\varepsilon_{g1}, \varepsilon_{g2}, ... \varepsilon_{gI_g}\right)'\right]$ be $\boldsymbol{\Omega}_g$. As usual, appropriate scale and level normalization must be imposed on $\boldsymbol{\Omega}_g$ for identification (more on this later). Under the utility maximization paradigm, $U_{gi_g} - U_{gm_g}$ must be less than zero for all $i_g \neq m_g$, since the individual chose alternative m_g. Let $y^*_{gi_g m_g} = U_{gi_g} - U_{gm_g} (i_g \neq m_g)$, and stack the latent utility differentials into a vector $\boldsymbol{y}^*_{\boldsymbol{g}} = \left[\left(y^*_{g1m_g}, y^*_{g2m_g}, ..., y^*_{gI_g m_g}\right)'; i_g \neq m_g\right]$. $\boldsymbol{y}^*_{\boldsymbol{g}}$ has a mean vector of $\boldsymbol{B_g} (\boldsymbol{\beta}_1' \boldsymbol{z}_{\boldsymbol{g1m_g}}, \boldsymbol{\beta}_1' \boldsymbol{z}_{\boldsymbol{g2m_g}}, ... \boldsymbol{\beta}_1' \boldsymbol{z}_{\boldsymbol{gI_g m_g}})'$, where $\boldsymbol{z}_{\boldsymbol{gi_g m_g}} = \boldsymbol{x}_{\boldsymbol{gi_g}} - \boldsymbol{x}_{\boldsymbol{gm_g}}, i_g = 1,2,..I_g; i_g \neq m_g$. To obtain the covariance matrix of $\boldsymbol{y}^*_{\boldsymbol{g}}$, define $\boldsymbol{M}_{\mathbf{g}}$ as an (I_g-1) x I_g matrix that corresponds to an (I_g-1) identity matrix with an extra column of -1's added as the m_g^{th} column. Then, one may write:

$$\boldsymbol{y}^*_{\boldsymbol{g}} \sim N\left(\boldsymbol{B}_{\boldsymbol{g}}, \boldsymbol{\Sigma}^*_{\mathbf{g}}\right), \text{where } \boldsymbol{\Sigma}^*_{\mathbf{g}} = \boldsymbol{M}_{\boldsymbol{g}} \boldsymbol{\Omega}_{\mathbf{g}} \boldsymbol{M}_{\boldsymbol{g}}'. \tag{2}$$

The discussion above focuses on a single nominal variable g. When there are G nominal variables, consider the stacked $\left[\sum_{g=1}^{G} (I_g - 1)\right] \times 1 -$vector $\boldsymbol{y}^* = \left[\left(\boldsymbol{y}_{\boldsymbol{1}}^{*'}, \boldsymbol{y}_{\boldsymbol{2}}^{*'}, ... \boldsymbol{y}_{\boldsymbol{G}}^{*'}\right)'\right]$, each of whose element vectors is formed by differencing utilities of alternatives from the chosen alternative m_g for the g^{th} nominal variable. Next, one may write:

$\boldsymbol{y}^* \sim N(\boldsymbol{B}, \boldsymbol{\Sigma}^*)$, where $\boldsymbol{B} = \left(\boldsymbol{B}_1', \boldsymbol{B}_2', .. \boldsymbol{B}_G'\right)'$ and $\boldsymbol{\Sigma}^*$ is a $\left[\sum_{g=1}^{G}(I_g - 1)\right] * \left[\sum_{g=1}^{G}(I_g - 1)\right]$ matrix as follows:

$$\boldsymbol{\Sigma}^* = \begin{bmatrix} \boldsymbol{\Sigma}_1^* & \boldsymbol{\Sigma}_{12}^* & . & . & . & \boldsymbol{\Sigma}_{1G}^* \\ \boldsymbol{\Sigma}_{21}^* & \boldsymbol{\Sigma}_2^* & . & . & . & \boldsymbol{\Sigma}_{2G}^* \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \boldsymbol{\Sigma}_{G1}^* & \boldsymbol{\Sigma}_{G2}^* & . & . & . & \boldsymbol{\Sigma}_G^* \end{bmatrix} \tag{3}$$

The off-diagonal elements in $\boldsymbol{\Sigma}^*$ capture the dependencies across the utility differentials of different nominal variables, the differential being taken with respect to the chosen alternative for each nominal variable.

Now, assume that, in addition to the G nominal variables, there are H continuous variables $(y_1, y_2, ...y_H)$ with an associated index h (h = 1, 2,…H). Let $y_h = \boldsymbol{\gamma}_h' \boldsymbol{s}_h + \eta_h$ in the usual linear regression fashion. Stacking the H continuous variables into a (H x 1)-vector $\boldsymbol{y}$, one may write $y = N(\boldsymbol{c}, \boldsymbol{\Sigma})$, where $\boldsymbol{c} = \left(\boldsymbol{\gamma}_1' \boldsymbol{s}_1, \boldsymbol{\gamma}_2' \boldsymbol{s}_2, ... \boldsymbol{\gamma}_H' \boldsymbol{s}_H\right)'$, and $\boldsymbol{\Sigma}$ is the covariance matrix of $\boldsymbol{\eta} = (\eta_1, \eta_2, ...\eta_H)$. The variance of $\tilde{\boldsymbol{y}} = \left(\boldsymbol{y}^{*'}, \boldsymbol{y}'\right)'$ can be written as:

$$\text{Var}(\tilde{\boldsymbol{y}}) = \tilde{\boldsymbol{\Lambda}} = \begin{bmatrix} \boldsymbol{\Sigma}^* & \boldsymbol{\Sigma}_{\mathbf{y}^*\mathbf{y}} \\ \boldsymbol{\Sigma}_{\mathbf{y}^*\mathbf{y}}' & \boldsymbol{\Sigma} \end{bmatrix}, \tag{4}$$

where $\boldsymbol{\Sigma}_{\mathbf{y}^*\mathbf{y}}$ is a $\left[\sum_{g=1}^{G}(I_g - 1)\right]xH$ matrix capturing covariance effects between the $\boldsymbol{y}^*$ vector and the $\boldsymbol{y}$ vector. The conditional distribution of $\boldsymbol{y}^*$, given $\boldsymbol{y}$, is multivariate normal with mean $\tilde{\boldsymbol{B}} = \boldsymbol{B} + \boldsymbol{\Sigma}_{\mathbf{y}^*\mathbf{y}} \boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{c})$ and variance $\tilde{\boldsymbol{\Sigma}}^* = \boldsymbol{\Sigma}^* - \boldsymbol{\Sigma}_{\mathbf{y}^*\mathbf{y}} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{\mathbf{y}^*\mathbf{y}}'$. The basis for the construction of the $\tilde{\boldsymbol{\Lambda}}$ matrix will be different for different individuals, since the chosen alternative for each nominal variable will, in general, be different across individuals. At the same time, it must be ensured that $\tilde{\mathbf{\Lambda}}$ across individuals is derived from a common covariance matrix $\mathbf{\Lambda}$ for the original $\left[\left(\sum_{g=1}^{G} I_g\right) + H\right]$-error term vector $(\boldsymbol{\varepsilon}', \eta')'$, subject to identification considerations $[\varepsilon = \left(\varepsilon_1', \varepsilon_2', .. \varepsilon_G'\right)']$. Also, the overall matrix $\tilde{\mathbf{\Lambda}}$ needs to be positive definite (as will be discussed later).[9]

Next, let $\boldsymbol{\theta}$ be the collection of parameters to be estimated: $\boldsymbol{\theta} = \left\lfloor \boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \ldots \boldsymbol{\beta_G}; Vech(\mathbf{\Sigma}^*); \boldsymbol{\gamma_1}, \boldsymbol{\gamma_2}, \ldots \boldsymbol{\gamma_H}; Vech(\mathbf{\Sigma}); Vech(\mathbf{\Sigma}_{\mathrm{y}^*\mathrm{y}}) \right\rfloor$, where Vech($\mathbf{\Sigma}$) represents the vector of upper triangle elements of $\mathbf{\Sigma}$. Then the likelihood function for the individual may be written as:

$$L\,(\boldsymbol{\theta}) = \phi_H(\boldsymbol{y} - \boldsymbol{c} \mid \mathbf{\Sigma}) \times F_{\tilde{G}}\left\lfloor -\tilde{\boldsymbol{B}}, \tilde{\mathbf{\Sigma}}^* \right\rfloor, \tag{5}$$

where $\phi_H(. \mid \mathbf{\Sigma})$ is the H-dimensional normal density with mean 0 and covariance matrix $\mathbf{\Sigma}$, and $F_{\tilde{G}}(.,.)$ is the $\tilde{G} = \left(\left[\sum_{g=1}^{G}(I_g - 1)\right]\right)$-dimensional normal cumulative distribution function.

The above likelihood function involves the evaluation of a $\tilde{G}$-dimensional integral for each individual, which can be very expensive if there are several nominal variables or if each nominal variable can take a large number of values or a combination of the two. So, the Maximum Approximated Composite Marginal Likelihood (MACML) approach of Bhat (2011), in which the likelihood function only involves the computation of univariate and bivariate cumulative distributive functions, is used in this study.

[9] Note that if $\tilde{\mathbf{\Lambda}}$ is positive definite, then it immediately implies that $\mathbf{\Sigma}^*$ (and each of $\mathbf{\Sigma}_1^*, \mathbf{\Sigma}_2^*, \ldots, \mathbf{\Sigma}_G^*$) as well as $\mathbf{\Sigma}$ are all positive definite because of the property that any principal square sub-matrix of a positive definite matrix is also positive definite.

6.2.2 The MACML Estimation Approach

Consider the following (pairwise) composite marginal likelihood function formed by taking the products (across the G nominal variables) of the joint pairwise probability of the chosen alternatives m_g and m_l for the g^{th} and l^{th} nominal variables for an individual.

$$L_{CML}(\boldsymbol{\theta}) = \phi_H(\boldsymbol{y} - \boldsymbol{c} \mid \boldsymbol{\Sigma}) * \prod_{g=1}^{G-1} \prod_{l=g+1}^{G} \Pr(d_{i_g} = m_g, d_{i_l} = m_l), \tag{6}$$

where d_{i_g} is an index for the individual's choice for the g^{th} nominal variable, and m_g is the actual chosen alternative for the g^{th} nominal variable. One can write:

$$\Pr(d_{i_g} = m_g, d_{i_l} = m_l) = F_{\tilde{G}_{gl}}(\boldsymbol{\Delta_{gl}}\tilde{\boldsymbol{B}}, \boldsymbol{\Delta_{gl}}\tilde{\boldsymbol{\Sigma}}^{*}\boldsymbol{\Delta}'_{\mathbf{gl}}), \tag{7}$$

where $\tilde{G}_{gl} = I_g + I_l - 2$ (I_g is the number of alternatives for the g^{th} nominal variable) and $\boldsymbol{\Delta_{gl}}$ is a $\tilde{G}_{gl} * \tilde{G}$ -selection matrix with an identity matrix of size ($I_g - 1$) occupying the first ($I_g - 1$) rows and the $\left[\sum_{j=1}^{g-1}(I_j - 1) + 1\right]^{th}$ through $\left[\sum_{j=1}^{g}(I_j - 1)\right]^{th}$ columns (with the convention that $\sum_{j=1}^{0}(I_j - 1) = 0$, and another identity matrix of size ($I_l - 1$) occupying the last ($I_l - 1$) rows and the $\left[\sum_{j=1}^{l-1}(I_j - 1) + 1\right]^{th}$ through $\left[\sum_{j=1}^{l}(I_j - 1)\right]^{th}$ columns. The net result is that the pairwise likelihood function now only needs the evaluation of a $\tilde{G}_{gl}$-dimensional cumulative normal distribution function (rather than the $\tilde{G}$ -dimensional cumulative distribution function in the maximum likelihood function). This can lead to substantial computation efficiency. However, in cases where there are several alternatives for one or more nominal variables, the dimension $\tilde{G}_{gl}$ can still be quite high. This is where the use of an analytic approximation of the multivariate normal cumulative distribution (MVNCD) function, as shown in Bhat (2011), is convenient. The resulting maximum approximated composite marginal likelihood (MACML) of Bhat (2011), which combines the CML approach with the analytic approximation for the MVNCD function evaluation, is solely based on bivariate and univariate cumulative normal computations. The MACML approach can be applied using a simple optimization approach for likelihood estimation. It also represents a conceptually simpler alternative to simulation techniques. Also, the MACML estimator $\hat{\boldsymbol{\theta}}_{MACML}$ is asymptotically normal distributed with mean $\boldsymbol{\theta}$ and covariance matrix given by the inverse of the Godambe's (1960) sandwich information matrix $G(\boldsymbol{\theta})$ (Zhao and Joe, 2005):

$$V_{MACML}(\boldsymbol{\theta}) = [G(\boldsymbol{\theta})]^{-1} = H(\boldsymbol{\theta})[J(\boldsymbol{\theta})]^{-1}[H(\boldsymbol{\theta})], \tag{8}$$

where $H(\boldsymbol{\theta})$ and $J(\boldsymbol{\theta})$ take the following form:

$$H(\boldsymbol{\theta}) = E\left[-\frac{\partial^2 \log L_{MACML}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right] \text{ and } J(\boldsymbol{\theta}) = E\left[\left(\frac{\partial \log L_{MACML}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial \log L_{MACML}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right)\right]$$

$H(\boldsymbol{\theta})$ and $J(\boldsymbol{\theta})$ can be estimated in a straightforward manner at the MACML estimate $\hat{\boldsymbol{\theta}}_{\boldsymbol{MACML}}$ as follows (introducing q as the index for individuals):

$$\begin{aligned} \hat{H}(\hat{\boldsymbol{\theta}}) &= -\left[\sum_{q=1}^{Q} \frac{\partial^2 \log L_{MACML,q}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]_{\hat{\theta}_{CML}}, \text{ and} \\ \hat{J}(\hat{\boldsymbol{\theta}}) &= \sum_{q=1}^{Q} \left[\left(\frac{\partial \log L_{MACML,q}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial \log L_{MACML,q}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right)\right]_{\hat{\boldsymbol{\theta}}_{\boldsymbol{CML}}}. \end{aligned} \tag{9}$$

6.2.3 Ensuring Identification and Positive Definiteness

There are two important issues that need to be dealt with during estimation, each of which is discussed in this section.

<u>Identification</u>

The estimated model needs to be theoretically identified. As discussed earlier, in a model with a nominal dependent variable, only utility differences matter. Suppose one considers utility differences with respect to the first alternative for each of the G nominal variables. Then, the analyst can restrict the variance term of the top left diagonal of the resulting covariance matrix (say $\breve{\boldsymbol{\Sigma}}_g^*$) of utility differences to 1 to account for scale invariance.

However, note that the matrix $\breve{\mathbf{\Sigma}}_g^*$ is different from the matrix $\mathbf{\Sigma}_g^*$ which corresponds to the covariance of utility differences taken with respect to the chosen alternative for the individual. Next, create a matrix of dimension $\left[\sum_{g=1}^{G}(I_g-1)\right]\times\left[\sum_{g=1}^{G}(I_g-1)\right]$ similar to that of $\mathbf{\Sigma}^*$ in Equation (3), except that the matrix is expressed in terms of utility differences with respect to the first alternative for each nominal variable:

$$\breve{\mathbf{\Sigma}}^* = \begin{bmatrix} \breve{\mathbf{\Sigma}}_1^* & \breve{\mathbf{\Sigma}}_{12}^* & \cdot & \cdot & \cdot & \breve{\mathbf{\Sigma}}_{1G}^* \\ \breve{\mathbf{\Sigma}}_{21}^* & \breve{\mathbf{\Sigma}}_2^* & \cdot & \cdot & \cdot & \breve{\mathbf{\Sigma}}_{2G}^* \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \breve{\mathbf{\Sigma}}_{G1}^* & \breve{\mathbf{\Sigma}}_{G2}^* & \cdot & \cdot & \cdot & \breve{\mathbf{\Sigma}}_G^* \end{bmatrix} \tag{10}$$

Further, construct an enhanced covariance matrix that includes the covariance matrix $\mathbf{\Sigma}$ of $\boldsymbol{\eta}=(\eta_1,\eta_2,....\eta_H)$ as follows:

$$\breve{\mathbf{\Omega}} = \begin{bmatrix} \breve{\mathbf{\Sigma}}^* & \breve{\mathbf{\Psi}} \\ \breve{\mathbf{\Psi}}' & \mathbf{\Sigma} \end{bmatrix}, \tag{11}$$

where the $\left[\sum_{g=1}^{G}(I_g-1)\right]\times H-$matrix $\breve{\mathbf{\Psi}}$ contains the covariances between the latent utility differentials (taken with respect to the first alternative) and the $\boldsymbol{y}$ vector. All elements of the matrix $\breve{\mathbf{\Omega}}$ are identifiable, and are the ones estimated. In the general case, this allows the estimation of $\sum_{g=1}^{G}\left(\frac{I_g*(I_g-1)}{2}-1\right)$ terms across all the G nominal variables (originating from $\left(\frac{I_g*(I_g-1)}{2}-1\right)$ terms embedded in each $\breve{\mathbf{\Sigma}}_g^*$ matrix; g=1,2,…G), $\sum_{g=1}^{G-1}\sum_{l=g+1}^{G}(I_g-1)\times(I_l-1)$ covariance terms in the off-diagonal matrices of the $\breve{\mathbf{\Sigma}}^*$ matrix characterizing the dependence between the latent utility differentials (with respect to the first alternative) across the nominal variables (originating from $(I_g - 1) \times (I_l - 1)$ estimable covariance terms within each off-diagonal matrix $\breve{\boldsymbol{\Sigma}}^{*}_{\mathbf{gl}}$ in $\breve{\boldsymbol{\Sigma}}^{*}$), $\left[\sum_{g=1}^{G}(I_g - 1)\right] \times H$ covariance terms in $\breve{\boldsymbol{\Psi}}$ for the dependence between the latent utility differentials and the linear regression errors, and the $[H \times (H+1)]/2]$ covariance terms in $\boldsymbol{\Sigma}$.

To construct the general covariance matrix $\boldsymbol{\Lambda}$ for the original $\left[\left(\sum_{g=1}^{G} I_g\right) + H\right]$-error term vector $(\boldsymbol{\varepsilon}', \boldsymbol{\eta}')'$, while also ensuring all parameters are identifiable, zero row and column vectors are inserted for the first alternatives of each nominal variable in $\breve{\boldsymbol{\Omega}}$. To do so, define a matrix $\boldsymbol{D}$ of size $\left[\left(\sum_{g=1}^{G} I_g\right) + H\right] \times \left[\left(\sum_{g=1}^{G}(I_g - 1)\right) + H\right]$. The first I_1 rows and $(I_1 - 1)$ columns correspond to the first nominal variable. Insert an identity matrix of size $(I_1 - 1)$ after supplementing with a first row of zeros into this first I_1 rows and $(I_1 - 1)$ columns of $\boldsymbol{D}$. The rest of the columns for the first I_1 rows and the rest of the rows for the first $(I_1 - 1)$ columns take a value of zero. Next, rows $(I_1 + 1)$ through $(I_1 + I_2)$ and columns (I_1) through $(I_1 + I_2 - 2)$ correspond to the second nominal variable. Again position an identity matrix of size $(I_2 - 1)$ after supplementing with a first row of zeros into this position. Continue this for all G nominal variables. Finally, insert an identity matrix of size H into the last H rows and H columns of the matrix $\boldsymbol{D}$ (with all other columns of these last H rows and all other rows of these last H columns receiving a value of zero). Thus, for the case with two nominal variables, one nominal variable with 3 alternatives and the second with four alternatives, and two continuous variables, the matrix $\boldsymbol{D}$ takes the form shown below:

$$\left[\begin{array}{cc|ccc|cc} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right]_{9*7}$$

Then, the general covariance matrix may be developed as $\mathbf{\Lambda} = \mathbf{D}\breve{\mathbf{\Omega}}\mathbf{D}'$. All parameters in this matrix are identifiable by virtue of the way this matrix is constructed based on utility differences and, at the same time, it provides a consistent means to obtain the covariance matrix $\tilde{\mathbf{\Lambda}}$ that is needed for estimation (and is with respect to each individual's chosen alternative for each nominal variable). Specifically, define a matrix ***M*** of size $\left[\left(\sum_{g=1}^{G}(I_g - 1)\right) + H\right] \times \left[\left(\sum_{g=1}^{G} I_g\right) + H\right]$. The first $(I_1 - 1)$ rows and I_1 columns correspond to the first nominal variable. Insert an identity matrix of size $(I_1 - 1)$ after supplementing with a column of '-1' values in the column corresponding to the chosen alternative. The rest of the columns for the first $(I_1 - 1)$ rows and the rest of the rows for the first I_1 columns take a value of zero. Next, rows (I_1) through $(I_1 + I_2 - 2)$ and columns $(I_1 + 1)$ through $(I_1 + I_2)$ correspond to the second nominal variable. Again position an identity matrix of size $(I_2 - 1)$ after supplementing with a column of '-1' values in the column corresponding to the chosen alternative. Continue this procedure for all *G* nominal variables. Finally, insert an identity matrix of size *H* into the last *H* rows and *H* columns of the matrix ***M***. With the matrix ***M*** as defined, the covariance matrix $\tilde{\mathbf{\Lambda}}$ for any individual is given by $\tilde{\mathbf{\Lambda}} = \mathbf{M}\mathbf{\Lambda}\mathbf{M}'$.

Positive Definiteness

The matrix $\tilde{\boldsymbol{\Lambda}}$ for any individual has to be positive definite. The simplest way to guarantee this is to ensure that the matrix $\breve{\boldsymbol{\Omega}}$ is positive definite (recall that this is the covariance matrix for the utility differentials with respect to the first alternative). To do so, the Cholesky matrix of $\breve{\boldsymbol{\Omega}}$ may be used as the matrix of parameters to be estimated. However, note that the top diagonal element of each $\breve{\boldsymbol{\Sigma}}_g^*$ is normalized to one for identification, and this restriction should be recognized when using the Choleski factor of $\breve{\boldsymbol{\Omega}}$. This can be achieved by appropriately parameterizing the diagonal elements of the Cholesky decomposition matrix. Thus, consider the lower triangular Choleski matrix $\breve{\boldsymbol{L}}$ of the same size as $\breve{\boldsymbol{\Omega}}$. Whenever a diagonal element (say the kk^{th} element) of $\breve{\boldsymbol{\Omega}}$ is to be normalized to one, the first element in the corresponding row of $\breve{\boldsymbol{L}}$ is written as $\sqrt{1-\sum_{j=2}^{k} d_{kj}^2}$, where the d_{kj} elements are the Cholesky factors that are to be estimated. With this parameterization, $\breve{\boldsymbol{\Omega}}$ obtained as $\breve{\boldsymbol{L}}\breve{\boldsymbol{L}}'$ is positive definite and adheres to the scaling conditions. Using this, one constructs $\boldsymbol{\Lambda}$, and subsequently obtains $\tilde{\boldsymbol{\Lambda}}$ as discussed in the previous section. The resulting $\tilde{\boldsymbol{\Lambda}}$ is positive definite, since it is constructed to be consistent with $\breve{\boldsymbol{\Omega}}$, which is positive-definite.

6.3 Data Description

The data for this study is derived from the 2009 National Household Travel Survey of the United States. This survey collects detailed socio-economic, demographic, travel, and vehicle information for a sample of households in the nation. Each trip (involving a personal automobile use) is tagged with the identity of the vehicle in the household that was used for the trip. Trip chains or tours can be easily constructed from the trip file. For this study, all closed loops or chains that began and ended at home were constructed as home-based tours and those that began and ended at work were constructed as work-based tours. As the analysis involves the choice of vehicle type, only tours undertaken by individuals in households that have multiple vehicles were chosen for analysis. Presumably, individuals in households with zero or one vehicle do not have a choice in vehicle usage. In addition, as vehicle type choice is likely to be limited at the home anchor, only home-based tours were selected for inclusion in the analysis sample. As tours involving journey to and from work are often time-space constrained and may involve aspects that constrain vehicle type choice (e.g., service workers who need pick-up truck for transporting tools of the trade), only home-based non-work tours were considered for analysis. Finally, the analysis was limited to home-based non-work vehicle tours undertaken on regular weekdays – Monday through Thursday – by individuals aged 15 years or over. These filtering criteria resulted in a total sample size of 66,030 home-based non-work tours suitable for analysis. For ease of computation, and to avoid the artificial inflation of test statistics that may lead to erroneous inferences, a random sample of 6,478 tours (nearly 10 percent) were selected for model estimation. Table 6.1 provides descriptive statistics for the subsample of HBNW tours. Each HBNW tour involved an average of 1.7 stops with average travel duration of 37 minutes and average tour length of 15.7 miles. On average, there were about 1.7 persons on each tour, reflecting the higher vehicle occupancies often associated with non-work travel. Each household in the subsample comprised of nearly three persons with one child. Most of the households in the sample (68 percent) reside in urban areas. There is a slightly higher percentage (54 percent) of females than males. This may be an artifact of limiting the analysis to non-work tours (e.g., involving household maintenance, serve-child) which may be undertaken more so by women than men. As the analysis is limited to tours undertaken by individuals in multi-vehicle households, the average vehicle ownership for the analysis sample is quite high at 2.8 cars per household. Nearly 20 percent of households report having four or more cars, reasonably consistent with the fact that the sample has 34 percent of households with four or more persons.

Table 6.1 also shows the distribution of vehicle types chosen for the tours in the estimation sample. First, the distribution is chosen for all tours. It is found that 42 percent of all tours are undertaken by auto, 25 percent by SUV, 14 percent by van, and 19 percent by pick-up truck. While these percentages might suggest that individuals are more inclined to choose cars and SUVs for travel, that may not necessarily be true because these percentages do not account for the differential availability of different vehicle types in the fleet. When one controls for vehicle availability in the fleet, then it is found that auto, van, and SUV all enjoy virtually identical probabilities of being chosen at about 50 percent. Only the pick-up truck has a lower probability of being chosen at about 32 percent. In other words, when auto, van, or SUV is available in the household fleet, each of these vehicles has a one-in-two chance of being chosen for a tour. When a pick-up truck is available in the fleet, the probability of its being selected for a tour is only about one-in-three. The percent of all tours undertaken by auto is greater than that for all other vehicle types simply because it is more available (present) in the household fleets. This demonstrates the importance of accounting for differential choice set composition when estimating models of vehicle type choice and drawing inferences regarding vehicle choices. Additional detailed statistics on tour attributes by vehicle type are shown in Table 6.2. The table is rather self-explanatory with descriptive statistics consistent with expectations. Average vehicle occupancy, for example, is greater for tours undertaken by van and SUV, presumably because these vehicles are likely to be owned and used by larger size households.

Table 6.1 Descriptive Statistics

Variable	Statistic
Tour-level	Mean
Number of passengers on the tour	1.7
Number of trips on the tour	2.7
Number of stops	1.7
Travel duration of the tour	37.0 minutes
Travel length of the tour	15.7 miles
Vehicle type chosen (all tours)	
% Auto	41.9
% Van	14.1
% SUV	25.4
% Pick-up Truck	18.6
Vehicle type chosen (all tours; accounting for vehicle availability in fleet)	
% Auto	51.1
% Van	50.2
% SUV	50.9
% Pick-up Truck	32.3
Household-level	
Average household size	3.1
% 1 person household	3.1
% 2 person household	44.5
% 3 person household	18.4
% 4+ person household	34.0
Average household vehicle count	2.8
% 2 vehicle household	49.9
% 3 vehicle household	30.7
% 4 or more vehicle household	19.4
Household Income	
% households with income < $40K	22.0
% households with income ≥40K and < 60K	17.3
% households with income ≥60K and < 100K	27.8
% households with income ≥100K	27.0
Average number of adults	2.3
Ratio of household size to vehicle count	1.2
Ratio of number of children to number of drivers	0.3
% households in non-urban area	32.0
% households that own the housing unit	95.2
Person-level	
Average age	52.2
% people ≥15 and < 25 years	9.0
% people ≥25 and < 45 years	22.3
% people ≥45 and < 65 years	41.6
% people ≥65 years	27.1
% males	46.0
% Hispanic respondents	6.4
% part-time employees	15.3
% full-time employees	29.2

Table 6.2 Tour Characteristics by Vehicle Type Chosen and Vehicle Fleet

Vehicle Body Type Selected for the Tour	Vehicle Fleet by Body Type	Frequency	Tour Distance	Travel Time	Number of passengers	Number of Stops
Average Tour Distance (Not Considering Vehicle Fleet Composition)						
Car		2716	16.0	37.7	1.6	1.6
Van		911	15.2	37.0	2.1	1.8
SUV		1647	15.4	36.0	1.8	1.7
Pickup		1204	15.6	36.5	1.5	1.6
Average Tour Distance (Considering Vehicle Fleet Composition)						
Car	Car, Pickup	1204	17.1	39.3	1.6	1.7
Car	Car, SUV	767	14.3	36.1	1.5	1.6
Car	Car, SUV, Pickup	196	16.6	36.9	1.6	1.6
Car	Car, Van	392	15.2	36.5	1.7	1.6
Car	Car, Van, Pickup	99	15.8	35.6	1.6	1.5
Car	Car, Van, SUV	47	17.2	41.2	1.5	1.6
Car	Car, Van, SUV, Pickup	11	20.3	42.4	1.6	1.6
Van	Car, Van	450	15.0	37.5	2.1	1.8
Van	Car, Van, Pickup	102	17.0	38.9	2.1	1.9
Van	Car, Van, SUV	50	11.1	28.8	1.8	1.6
Van	Car, Van, SUV, Pickup	12	21.7	43.8	2.3	1.8
Van	Van, Pickup	169	14.4	35.9	2.0	1.8
Van	Van, SUV	100	17.2	39.6	2.1	1.7
Van	Van, SUV, Pickup	28	15.4	31.9	1.9	1.3
SUV	Car, SUV	824	14.1	34.3	1.7	1.6
SUV	Car, SUV, Pickup	241	16.0	37.4	1.7	1.7
SUV	Car, Van, SUV	46	15.4	36.4	1.5	1.9
SUV	Car, Van, SUV, Pickup	17	18.3	40.4	2.1	1.9
SUV	SUV, Pickup	412	17.0	37.4	1.9	1.8
SUV	Van, SUV	76	16.4	40.3	1.7	1.7
SUV	Van, SUV, Pickup	31	17.9	38.4	1.9	1.6
Pickup	Car, Pickup	662	15.4	35.8	1.5	1.6
Pickup	Car, SUV, Pickup	137	16.9	36.9	1.4	1.6
Pickup	Car, Van, Pickup	51	14.0	33.5	1.6	1.4
Pickup	Car, Van, SUV, Pickup	10	15.0	33.6	1.4	1.6
Pickup	SUV, Pickup	221	15.6	37.4	1.5	1.6
Pickup	Van, Pickup	111	15.9	37.5	1.4	1.7
Pickup	Van, SUV, Pickup	12	17.4	61.4	1.3	1.3

All figures are averages except for the frequency column. Distance is in miles and travel time is in minutes.

6.4 Model Estimation Results

This section presents a detailed discussion of the model estimation results. A variety of models were estimated to understand the nature of relationships among the four tour attributes considered in this research.

6.4.1 Structure of Relationships among Endogenous Variables

Before proceeding to a discussion of the estimates of coefficients and error covariances, it may be beneficial to consider behavioral hypotheses governing the nature of relationships among the endogenous variables. Figure 6.1 presents a flow chart depicting the structure of relationships that guided the model specification and estimation. Socio-economic and demographic attributes are assumed to affect all endogenous variables. Among the endogenous variables themselves, passenger accompaniment (which is an endogenous variable because it is a function of explanatory variables) is assumed to impact tour complexity. In the context of a joint or partly joint tour, it is more likely that additional stops will be made to serve the needs of the passenger or to engage in a series of joint activities (for example, eat dinner at a restaurant and then go to the movies). In addition, however, passenger accompaniment may also affect vehicle type choice. When there are multiple individuals involved in a trip, then the larger vehicle may be chosen for reasons of comfort. Finally, passenger accompaniment may also affect tour length. When a joint activity is involved, or a passenger needs to be dropped off or picked up, then destinations are often dictated by the collective needs and desires of the multiple occupants. This may result in traveling to and from locations that are farther away than would otherwise be the case. As such, passenger accompaniment is postulated as affecting all three of the other endogenous variables.

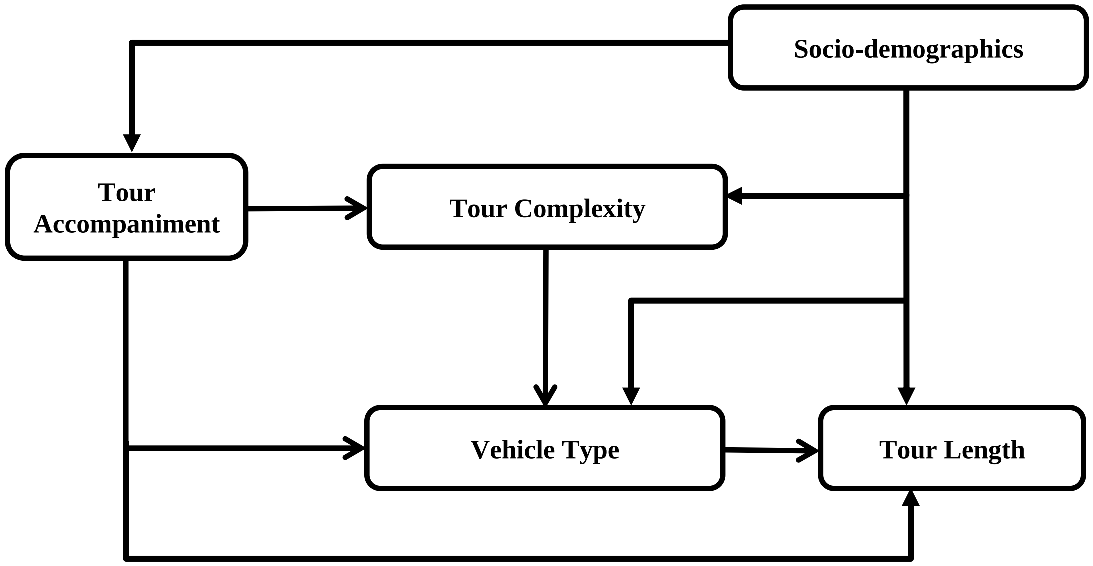

Figure 6.1 Framework of Relationships among Endogenous Variables

Next, consider tour complexity which is a binary choice variable indicating whether the tour involved a single stop or multiple stops. Tour complexity is assumed to impact both vehicle type choice and tour length. In the context of vehicle type choice, it is possible that larger and more comfortable vehicle types will be used for multiple stop tours. Also, multi-stop tours are likely to be of longer distance because of the need to travel to multiple locations. As multi-stop tours are longer in distance, two possibilities arise. Tour length may, in turn, influence vehicle type choice. First, if tour length is longer (because the tour is complex), individuals may choose the more fuel efficient vehicle type to reduce travel costs associated with traveling long distances. On the other hand, if a person would like to increase comfort levels during a long multi-stop tour, then the larger vehicle type may be chosen to undertake the trip.

In other words, the relationship between the last two variables is subject to debate. While the flowchart shows vehicle type choice affecting tour length, it is entirely possible that tour length affects vehicle type choice. If vehicle type choice affects tour length, then one is implying that people make conscious choices regarding destinations (miles of travel) depending on the nature of the vehicle being used. If a person is using a small fuel efficient car, would the person visit farther destinations and travel more miles because it is possible to do so at lower cost than if a gas guzzling vehicle were used? Or would the person visit close-by destinations and reduce mileage because traveling long distances in the small fuel efficient vehicle is uncomfortable? Alternatively, if the traveler has to visit destinations farther away, then would the fuel efficient vehicle be chosen to keep costs down? Or would a large gas guzzling vehicle be used to maximize comfort levels on the tour? In a previous study, Konduri *et al.* (2011) found that a model in which vehicle type choice affects tour length is statistically superior to a model specification in which tour length affects vehicle type choice. While that finding is clear and intuitive, as vehicle type choice (and allocation of vehicles in a household to drivers) is likely to be a longer term decision relative to tour length choices, the study did not account for the possible endogeneity of passenger accompaniment and tour complexity. Both of these dimensions were treated as exogenous variables, potentially resulting in erroneous inferences regarding the direction of the relationship between vehicle type choice and tour length. This study offers the opportunity to further explore the nature of the relationship between these two variables while accounting for the endogeneity of passenger accompaniment and tour complexity.

Table 6.3 Model Estimation Results

Variable Description	Coeff	t-stat	Coeff	t-stat	Coeff	t-stat
Accompaniment (Base Alternative: Solo)	Partly Joint		Joint			
Constant	-0.9168	-12.51	-0.6350	-11.71		
Socio-economic Attributes						
Ratio of household size to number of vehicles	0.3876	8.24	0.4578	9.55		
Ratio of number of children and number of drivers	0.4953	9.84	0.1875	3.78		
Race of household respondent is Caucasian	-0.1687	-3.27				
Gender (Male = 1, Female = 0)	-0.3065	-7.62	-0.2683	-7.80		
Age 18 years and younger (Yes=1, No=0)	-0.3278	-3.65	0.1688	2.27		
Part-time employment indicator (Yes = 1, No = 0)	0.1252	2.66				
Income indicator: 0 - $40K (Yes = 1, No = 0)			0.1839	4.89		
HH in non-urban area (Yes=1, No=0)			0.0688	2.08		
Tour Complexity (Base Alternative: Simple)	Complex					
Constant	-0.0860	-1.09				
Tour Attributes						
Accompaniment type: Partly Joint (Yes=1, No=0)	-0.3091	-1.98				
Accompaniment type: Joint (Yes=1, No=0)	-0.3197	-1.75				
Socio-economic Attributes						
Can set or change work start time (Yes=1, No=0)	-0.1354	-2.56				
Full-time employment indicator (Yes = 1, No = 0)	-0.0879	-2.20				
Race of household respondent is Hispanic	-0.1326	-1.93				
Age 18 years and younger (Yes=1, No=0)	-0.1648	-2.06				
Gender (Male = 1, Female = 0)	-0.1647	-4.18				
Vehicle Type (Base Alternative: Pickup Truck)	Auto		Van		SUV	
Constant	1.0414	24.21	0.6839	8.34	1.0239	23.22
Tour Attributes						
Accompaniment type: Partly Joint (Yes=1, No=0)	-0.5405	-3.31	0.4617	4.61		
Accompaniment type: Joint (Yes=1, No=0)			0.5195	6.18		
Socio-economic Attributes						
Gender (Male = 1, Female = 0)	-1.0920	-22.43	-1.3309	-18.12	-1.1220	-20.44
Age 65 years and older (Yes=1, No=0)			0.1374	1.77		
No. of children in household			0.0823	3.06		
HH in non-urban area (Yes=1, No=0)			-0.1168	-1.74	-0.0970	-2.27
Tour Length						
Constant	1.4364	14.53				
Tour Attributes						
Accompaniment type: Joint (Yes=1, No=0)	0.6098	3.17				
Tour complexity: Complex tour (Yes=1, No=0)	1.4469	7.89				
Vehicle type: Auto (Yes=1, No=0)	0.0872	2.51				
Vehicle type: Van (Yes=1, No=0)	0.1545	2.81				
Vehicle type: SUV (Yes=1, No=0)	0.0454	1.12				
Socio-economic Attributes						
Ratio of household size to number of vehicles	-0.0843	-2.10				
No. of children in household	-0.0575	-3.29				
Gender (Male = 1, Female = 0)	0.1173	3.91				
Education level: Atleast some college (Yes=1, No=0)	0.0469	1.78				
Can set or change work start time (Yes=1, No=0)	-0.0680	-1.79				
HH in non-urban area (Yes=1, No=0)	0.4470	16.81				
Income indicator: 0 - $40K (Yes = 1, No = 0)	-0.0843	-2.62				
Age 65 years or over (Yes=1, No=0)	-0.0703	-2.34				

6.4.2 Model Results

Model estimation results are presented in Table 6.3. The constants in the model of passenger accompaniment suggest that partly joint tours are the least likely tour type (all other things being equal) and solo tours are the most likely. A host of socio-economic attributes impact tour accompaniment. Tours undertaken by individuals in households with larger household size (relative to number of vehicles) or larger number of children (relative to the number of drivers) are more likely to be joint tours than solo tours as evidenced by the positive coefficients on these variables. In particular, the presence of children appears to induce partly joint tours, a finding that is consistent with the notion that such households undertake serve-child tours where the child accompanies the parent for a part of the tour and the driver is alone for the remainder of the tour. Males are less likely to undertake joint tours, possibly because females are more likely to take care of household responsibilities and chauffeuring of children. Those 18 years and younger are likely to undertake joint tours, but less likely to undertake partly joint tours; this finding is consistent with expectations, considering that the sample is restricted to those of driving age. These individuals probably drive themselves in solo tours as opposed to needing a partly joint tour involving a pick-up/drop-off. They are, however, more likely to engage in full joint tours in consort with other household members. Part-time employment is associated with greater participation in partly joint tour; perhaps part-time employees are more able to undertake serve-passenger and serve-child activities on behalf of the household leading to a greater prevalence of these partly joint tours for this demographic.

In general, complex tours are less likely to occur than simple tours (all other things being equal) as evidenced by the negative constant for the complexity utility equation (although the coefficient is not statistically significant). A rather surprising finding is that joint and partly joint tours are less likely to be complex than solo tours. One would have expected these tour types, that involve multiple passengers, to be more complex. On the other hand, it is possible that this finding is quite intuitive. When multiple passengers are involved, then the driver or any one individual may not be able to undertake a series of activities on a tour that are of no interest or relevance to other passengers on the tour. The individuals on the joint tour are collectively going to a certain location, undertaking a joint activity, and then returning to base. Only the activity that is of interest and relevance to the entire group is visited. Those who work full time are less likely to engage in complex tours, presumably because of time constraints associated with full time employment. Younger individuals are less likely to undertake complex tours. Males are also less likely to undertake complex tours, suggesting that females have more complex activity-travel patterns as they shoulder a greater share of household responsibilities.

The vehicle type choice model is presented next within Table 6.3. The constants are all very significant, with the auto and SUV having the highest constants suggesting that these two vehicle types tend to get used more often than others when they are in the choice set. Van also has a positive coefficient suggesting that it is used more than the pick-up truck which is the base alternative. As expected, joint tours are more likely to be undertaken by van or the larger vehicle type, suggesting that the more comfortable (larger) vehicle type is chosen when multiple occupants are involved. In the case of a partly joint tour, the auto is least likely to be used among all vehicle types. Males are more likely to use the pick-up truck (when it is available in the household fleet) compared to all other vehicle types when they are present in the fleet. This finding is consistent with expectations and illustrative of the strong gender influence in the pick-up truck market. Those aged 65 years or over and those with children in the household are most likely to use van for tours. The older age group may enjoy the comfort and driving ease of a van, and may have more use for the van as they transport grand children or grown children. It is not surprising that the presence of children is associated with a positive impact on van use; households with child transport duties would likely enjoy the space and comfort of van for chauffeuring duties. Households in non-urban areas are more likely to use a car or a pick-up truck in comparison to a van and SUV. This is also consistent with expectations in that the van and SUV are probably not the most preferred vehicle types in rural areas. It is interesting to note that tour complexity does not directly enter the equation of vehicle type choice. It appears that tour complexity does not truly directly influence vehicle type choice; rather it is the accompaniment that influences vehicle type choice.

Finally, the model of tour length shows that accompaniment, complexity, and vehicle type affect tour length. In other words, tour length is affected by all other tour attributes. According to the model estimation results presented in the last part of Table 6.3, joint tours are likely to be of longer length. This is consistent with the notion that tours involving multiple people might be longer in distance in an effort to find destinations that satisfy the desires of all individuals on the tour. Similarly, tour complexity also adds significantly to tour length. As one adds stops to a tour, it is natural to expect tour length to increase as the addition of each stop entails some additional travel distance. Among the vehicle types, vans tend to have a longer tour length, presumably because vans are comfortable for transporting household members or undertaking joint activities. It is somewhat surprising to see that cars are next in line in terms of a positive impact on tour length, while sports utility vehicles and pick-up trucks show the lowest impact on tour length. This finding is a key sign that people are making a conscious trade-off in the distance traveled by different vehicle types. Both sports utility vehicles and pick-up trucks generally have the poorest fuel economy among all vehicle types. The model is indicating that both of these vehicle types are associated with the shortest tour lengths relative to car and van vehicle types (both of which tend to have better fuel economy). It appears that individuals are making conscious decisions involving trade-offs between travel cost and miles of travel. If a large gas guzzling vehicle is used, then the individual may attempt to consciously find locations that are closer in distance to reduce travel costs. Of course, such trade-offs can be exercised only in the context of non-work tours/travel.

The impacts of socio-economic and demographic attributes on tour length are in line with expectations. As household size (relative to number of vehicles) increases, the tour length decreases. This may be reflective of the vehicle availability constraints that the household has to deal with. In households where household size is large relative to number of vehicles, individuals who take the vehicle to undertake a tour may have to return quickly so that another household member can use the same vehicle. This compels travelers to undertake short tours and minimize travel time. The lower number of vehicles relative to household size may also be reflective of a lower income level; individuals in such households may purposefully undertake shorter distance tours to save on travel costs. As the number of children increases, individuals tend to make shorter tours. This is presumably due to two reasons. First, if the children are accompanying the tour maker, then the individual may choose to complete errands quickly by undertaking shorter tours in order to avoid taxing the patience of the children. Second, if the children are not accompanying the individual on the tour, then it might be necessary for the individual to quickly conclude the tour and return home to tend to the children. It is also possible that children have schedule constraints that compel the traveler to undertake shorter tours. Males tend to make longer tours suggesting that females make shorter tours visiting destinations more closely located to the home base. Those with higher education undertake longer tours, perhaps because they have higher income levels, or are more aware of desirable destinations for non-work activities. As expected, those in non-urban areas undertake longer tours; this is likely due to the lower levels of accessibility to destinations enjoyed by such households. Lower income individuals make shorter tours as do individuals 65 years of age and over. Older individuals may not be comfortable traveling long distances. Those with flexible work start time, and thus less rigid time-space constraints associated with work, are found to engage in short tour lengths. This is presumably because these individuals do not have to chain multiple activities into longer multi-stop tours in the quest for efficiency; instead, they can engage in a larger number of short tours. Indeed, the work time flexibility is negatively associated with complex tour formation.

6.4.3 Model Assessment

This section presents a brief assessment of the joint model estimated and presented in this study. The log likelihood of the final joint model accounting for all potential correlations is significantly better than that of the independent model where all dimensions are estimated separately. The log-likelihood value for the joint model is -23487.87 while that for the independent model ignoring error correlations is -23535.74. The likelihood ratio test statistic is found to be 95.75 with 12 degrees of freedom. This value is considerably greater than the critical χ^2 value of 21.03 at 12 degrees of freedom, suggesting that the joint model offers a statistically superior fit at a 0.05 level of significance. This finding of improved goodness-of-fit of the joint model is the first indication that there may significant error correlations that contribute to a poorer fit in the independent model where they are ignored.

Table 6.4 Error Covariance Matrix

Dimension	Partly Joint	Joint	Complex	Auto	Van	SUV	Tour Length
Partly Joint	1						
Joint	0.5	1					
Complex	0.3525 (4.50)	0.1914 (1.95)	1				
Auto	0.3787 (4.52)	0.1233 (3.97)	0.0469 (1.69)	1			
Van	0	0	0	0.5	1		
SUV	0.1679 (3.96)	0.2057 (5.81)	0	0.6896 (2.39)	0.5	1	
Tour Length	0	0.2030 (1.72)	0.4549 (3.10)	0	0.1278 (2.97)	0	0.9999 (15.59)

Values in parentheses are t-statistics. If no t-statistic is provided, it means that the covariance was fixed to the shown value.

The covariance matrix $\breve{\Omega}$ for the utility differentials[10] with respect to the first alternative is shown in Table 6.4. Only those parameters that are free to be estimated have t-statistics reported against them. All other parameters are fixed during estimation. It can be seen that there are significant error correlations across different nominal variables and the continuous variable even after including right hand side endogenous variables in the equations that comprise the joint model system. In other words, even after accounting for observed relationships among the tour attributes considered in this study, there are correlated unobserved factors affecting these attributes leading to the estimation of significant error correlations. The interpretation of the error correlations is that unobserved attributes that affect one dimension are correlated with unobserved attributes that affect another dimension. In this particular study, it is found that all significant error correlations are positive. For example, unobserved factors that contribute to partly joint or joint tours are positively correlated with unobserved factors that contribute to complex tour formation. Suppose a person is a fun-seeking individual who likes to socialize and visit friends. Then, this unobserved attribute of the individual is likely to positively influence both joint tour formation and complex tour formation. Such individuals are likely to enjoy traveling with others (friends) leading to the formation of joint tours. Such individuals are also likely to visit multiple places and undertake complex tours as they seek to engage in fun activities with friends. They may also have to go to multiple locations to pick up and drop off friends.

Similar interpretations may be applied to other significant error correlations. For example, an adventurous individual may be inclined to undertake complex tours and longer tours in search of destinations that meet the individual's activity preferences. The bottom line is that there are significant error correlations, possibly stemming from attitudes and preferences that make individuals likely to bundle certain choice options together, or built environment and accessibility measures that were not included in the

[10] The t-statistics reported in the table are with respect to the corresponding values in an independent model where there are 1ˢ along the diagonal and 0.5ˢ for all off-diagonal elements in each of the block diagonal matrices corresponding to each nominal variable and 0ˢ for rest of the elements. It can also be seen that parameters which are fixed during the estimation process do not have t-statistics reported along with them.

model specifications of this study. The inclusion of such attributes in the model specifications remains a future research exercise. Unobserved attributes that contribute to an individual choosing the car also positively contribute to the choice of the sports utility vehicle as evidenced by the positive error correlation between auto and SUV vehicle type choices. Unobserved attributes that contribute to joint or complex tour formation are positively correlated with unobserved attributes that contribute to longer tours.

It is interesting to note that there are some key differences in model results between the multi-dimensional choice model system presented here and the bivariate model system estimated on the same data set presented in earlier research (Konduri *et al.*, 2011). In the bivariate model system where accompaniment and complexity were treated as exogenous variables without adequate accounting for endogeneity and correlated unobserved attributes simultaneously impacting these additional dimensions, the tour complexity was found to positively impact choice of van in the vehicle type choice model. However, in the model estimated for this study, tour complexity was not statistically significant at all in any of the vehicle type choice utility equations. Also, in the previous research effort, the influence of accompaniment on tour complexity was never captured because these two variables were treated as independent variables. In the earlier bivariate model, the number of error correlations that could be estimated was considerably smaller because only two choice dimensions were considered as endogenous. Among the error correlations, only the one between van type choice and tour length was found to be statistically significant. Other relevant error correlations that were found to be significant in this work were not found statistically significant in that simpler bivariate model system. Moreover, the error correlation between van and tour length was found to be negative in that earlier bivariate model. In the multi-dimensional model of this study, this error correlation is found to be positive, suggesting that model parameter estimates and inferences are significantly impacted by the lack of proper accounting for endogeneity in multivariate modeling contexts. The finding in this study suggests that unobserved attributes contributing to longer tours (such as living in suburban locations with lower accessibility to destinations) also contribute to the choice of van as a vehicle type (as households in these locations tend to have larger household sizes with children and may desire to use the van to accommodate multiple individuals more comfortably).

Finally, if one were to compare the model estimation results against the original hypothesized structure of the nature of the relationships among these endogenous variables as depicted in Figure 6.1, it is seen that the relationships postulated in that figure are all significant except the one where tour complexity affects vehicle type choice. It appears that tour complexity itself does not directly affect vehicle type choice. Rather there are common unobserved attributes that simultaneously impact tour complexity and vehicle type choice as evidenced by the positive significant error correlation between tour complexity and car vehicle type choice. However, this covariance is rather weakly significant (with a t-statistic of 1.69) and no other tour complexity – vehicle type choice error covariance is significant. In other words, according to the model estimated in this study, the relationship between tour complexity and vehicle type choice is quite tenuous, a finding substantially different from the earlier study (Konduri *et al.*, 2011) where tour complexity was found to significantly directly impact (positively) the choice of van. However, one of the key similarities between the findings of the two studies is that, in both cases, model specifications where vehicle type choice significantly affected tour length were statistically superior to model specifications where tour length affected vehicle type choice. Thus, the notion that vehicle type choice is a longer term decision, where vehicles are broadly allocated to adults or drivers in a household as a higher level household decision, appears to hold true regardless of whether one considers accompaniment and tour complexity as exogenous to the system or endogenous to the system. However, the multidimensional choice model estimation results in this study point to the influence that accompaniment and tour complexity have on vehicle type choice in the context of a tour. In other words, although vehicle allocation to adults may be occurring as a longer-term higher-level decision process, conscious decisions regarding vehicle type choice and tour length are being made at the tour level depending on the nature of the tour (in terms of accompaniment and complexity).

6.5 Conclusions

This study presents a multi-dimensional choice model system of tour attributes with a view to better understand the complex inter-relationships that exist among various choice dimensions of interest in the context of tour- and activity-based travel model specification. The four dimensions considered are passenger accompaniment, tour complexity (measured by number of stops undertaken), vehicle type chosen, and tour length (distance). Modeling these choice dimensions independently of one another, without recognizing the potential presence of correlated unobserved attributes that simultaneously impact multiple dimensions, leads to a number of limitations that may result in erroneous behavioral inferences and travel forecasts. First, when endogeneity exists among multiple choice dimensions that are modeled independently of one another in a series of sequential models loosely strung together, resulting parameter estimates are biased and inconsistent. This can lead to erroneous impact assessments and scenario forecasts. Second, it is entirely possible that some choice dimensions are made as a package or bundle by individuals. In the context of a tour, it is conceivable that choices regarding passenger accompaniment, stop formation, vehicle type, and locations to be visited constitute a package of choices that are made together in a bundle. When that happens, there are bound to be unobserved attitudinal and lifestyle preference variables that inevitably impact multiple dimensions simultaneously. Thus, a model that ignores the bundling or packaging of choices will inevitably be limited in its representation of behavioral processes at play.

The current research makes two major contributions to the field. First, the study presents an econometric methodology for estimating multi-dimensional choice model systems that include a variety of dependent variable types and accommodate error covariances across multiple dimensions. The modeling methodology takes advantage of the latest advances in model formulation and estimation, and involves the use of novel estimation techniques that greatly reduce computational burden without compromising the efficiency (precision) of parameter estimates. Second, the study sheds considerable light on the nature of the empirical relationships among the four dimensions examined. There is much interest in understanding how multiple tour attributes are related to one another with a view to better inform the structure and specification of tour-based models. In addition, there is very limited evidence on tour-level vehicle type choice processes despite the obvious importance of this choice dimension in the ongoing debate regarding energy sustainability and greenhouse gas emission reduction.

A simultaneous equations model is estimated on a sample of over 6000 tours drawn from the 2009 National Household Travel Survey of the United States. It is found that vehicle type choice is highly dependent on vehicle availability by type, underscoring the need to consider variable choice set composition explicitly when modeling vehicle type choice. The model estimation results show that the dimensions considered in this study are all related to one another. Passenger accompaniment affects tour complexity, with tours involving passengers likely to be of less complexity involving just a single stop as opposed to multiple stops. Passenger accompaniment, but not tour complexity, affects vehicle type choice with joint tours most likely to be undertaken by van. Passenger accompaniment, tour complexity, and vehicle type choice are all found to affect tour length. Joint tours tend to be longer in distance, as do complex tours involving multiple stops. Van tours tend to be longest in length, followed by car tours. Tours by SUV and pick-up truck tend to be shorter in length than van and car tours. In other words it appears that tours undertaken by more fuel efficient vehicles are likely to be longer than tours undertaken by SUV and pick-up trucks. The results point to the possible conscious choices and decision on the part of travelers to choose locations and travel distances consistent with the fuel efficiency of the vehicle that they drive.

The model in which vehicle type choice affected tour length was found to offer superior statistical fit than the model in which tour length was allowed to affect vehicle type choice. Moreover, the statistical fit of the simultaneous equations model with error covariances was considerably superior to the fit of the independent equations model with error covariances restricted to zero. This finding suggests that there are correlated unobserved attributes simultaneously impact multiple tour dimensions calling for the increased deployment of models such as that presented in this study. Further research is needed to fully understand the nature of the unobserved attributes affecting these multiple tour dimensions, but these are likely to be personal attitudes and preferences, built environment attributes, and accessibility measures, besides other unobserved variables (such as time-space constraints, household constraints, personal constraints, and institutional constraints) not available in the data set. The modeling methodology presented in this research has the potential to offer dramatic breakthroughs in the ability of the profession to better capture and represent simultaneous choice processes at play.

From a policy perspective, the findings of the study suggest that the complex inter-relationships among tour choice dimensions make the analysis of policy impacts potentially more involved than one might have imagined. The findings suggest that the use of a more fuel efficient vehicle for a tour contributes to the choice of a longer tour length. In other words, although the driving of a fuel efficient vehicle may reduce energy consumption and emissions, the finding that it is driven longer distances suggests that the energy consumption and emissions reductions may not be as much as expected and the increase in vehicle miles of travel may actually contribute to greater levels of congestion on roadways. Policies aimed at encouraging the ownership and use of fuel efficient and clean vehicles may end up not providing the originally intended benefits. Another interesting finding is that the flexibility associated with work start time is contributing to the formation of single stop tours (less complexity) of shorter length. In other words, the loosening of time-space constraints imposed by rigid work schedules makes it possible for people to undertake less efficient activity-travel patterns that are characterized by a higher frequency of single stop tours. While an individual single stop tour is likely to be of shorter length than a complex tour, the fact that there are more of them (assuming no change in activity agenda itself) could result in an increase in overall mileage. Again, from a policy perspective, the potential benefits that would be expected from the implementation of a flexible work hours strategy may not be fully realized.

In summary, the study points to the need to further develop multi-dimensional choice models capable of reflecting the complex observed and unobserved inter-relationships among several behavioral dimensions of interest. Such models would be able to more accurately capture behavioral processes at play and offer more robust forecasts of possible consequences of policy actions. Although the econometric model system formulated and presented in this study may appear to be a rather complex statistical exercise, it offers the potential to move the profession a step closer to implementing more simultaneous equations model systems that recognize the package nature of activity-travel choices.

CHAPTER 7: An Integrated Model of Residential Location, Work Location, Vehicle Ownership, and Commute Tour Characteristics

The material in this chapter is drawn substantially from the following technical paper:

Paleti, R., C.R. Bhat, and R.M. Pendyala, (2012) An integrated model of residential location, work location, vehicle ownership, and commute tour characteristics. Technical Paper, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin.

As mentioned earlier in Chapter 2, this chapter of my dissertation attempts to overcome the limitations associated with previous work in the specification and estimation of multi-dimensional model systems of location and activity-travel choices by tying together six choice dimensions in a joint modeling framework namely, - residential location and workplace location choices (long term multinomial choice variables), commute distance (which is an outcome of residential location and workplace location choices and is a long term continuous variable), household vehicle ownership (medium term ordinal dependent variable), commute mode choice multinomial travel choice variable), and finally, number of stops made during the commute tour (short term ordinal dependent variable). These six variables are tied together in a temporal framework as shown in Figure 7.1a while recognizing the bundling of these choice dimensions associated with the jointness or simultaneity in decision-making. The remainder of this chapter is organized as follows. Section 7.1 offers a description of the data, while Section 7.2 presents the methodology in detail. Section 7.3 presents model estimation results, and lastly Section 7.4 offers some concluding thoughts.

7.1 Data

The data for this study is derived from the 2009 National Household Travel Survey (NHTS) which is conducted by the US Department of Transportation on a periodic basis to obtain information about the travel characteristics of the population. The 2009 NHTS data is ideally suited for a study of this nature because it provides information about residential location (census tract in which the household resides), work location of employed individuals in the household (census tracts of work locations), household vehicle ownership, and detailed travel choices for a 24 hour travel diary period. In addition, the data include detailed socio-economic and demographic data about respondent households.

For the current study, the survey subsample from the San Francisco Bay Area is extracted for analysis and model estimation purposes. This was done to limit the scope of the geographic region, deal with manageable sample sizes, and take advantage of secondary census data for the region (available from a previous study) that can be merged to the records of the NHTS. As the current study involves the modeling of work location (among other dimensions), the subsample extracted for this study includes only employed individuals who have a fixed work location outside of the home location and who have provided complete travel diary data that includes information on commute tours, mode choice, and stop-making behavior.

Census tract data for the San Francisco Bay Area was merged with the NHTS data records to help characterize household and workplace locations. Instead of using the classic definition of spatial unit choice (identified by census tract or traffic analysis zone), this study employs categories of land use density to characterize location choices. This helps make the definition of choice alternatives clear and manageable and more effectively captures the notion that people are looking for a built environment (land use density) that suits their mobility and lifestyle preferences. In other words, people are not choosing between tract A or B, but rather between a unit that offers a built environment of certain attributes versus another unit that offers a different built environment. Residence and workplace locations are categorized into four possible alternatives based on housing unit density (housing units per square mile).

After extensive data cleaning, the final estimation sample includes 1,480 employed individuals. Besides residence and work locations, a number of other dependent variables were constructed for this sample. The commute distance is simply a measure of separation between the residence and work locations as reported in the travel diary. Vehicle ownership is reported by respondents as well. For commute tour mode, the mode that was used in the work-to-home (half) tour was designated as the chosen alternative. If transit was used for any leg of the journey, then the commute tour mode was designated as transit. Four modal alternatives – drive alone, shared ride, transit, and walk/bike – characterized the mode choice for more than 99 percent of the tours. The few people whose commute tours did not fall within one of these four modal alternatives were omitted from the final estimation sample. Finally, the total number of stops made during the home-to-work and work-to-home tours constituted the last dependent variable of the study.

The sample of 1,480 employed individuals exhibited socio-economic and demographic characteristics suitable for undertaking a model estimation effort such as that undertaken in this study. The distribution of individuals in the four residential location alternatives is as follows:

- 0-499 housing units per square mile: 22.6%
- 500-1999 housing units per square mile: 30.9%
- 2000-3999 housing units per square mile: 29.9%
- ≥ 4000 housing units per square mile: 16.6%

The distribution of individuals with respect to work locations is somewhat similar except that higher percent of individuals (32.4%) work in low density (0-499) tracts while a smaller percent (20.5%) of individuals work in higher density (2000-3999) tracts. With respect to vehicle ownership, 1.8 percent of the employed individuals indicate residing in households with no vehicle. This fraction is lower than that for the general population, but such differences are expected when considering a pure worker sample. About 47 percent of individuals reside in two-vehicle households, 23.2 percent reside in three-vehicle households, and 15 percent reside in households with four or more vehicles.

An examination of commute mode share shows that 72.6 percent of individuals commute by drive alone, 16.1 percent by shared ride, 8 percent by transit, and 3.2 percent by bicycle/walk. The average commute distance is 13.5 miles with a standard deviation of 14.4 miles. The distribution of stop-making shows that 47 percent of commuters make zero (non-work) stops within the commute tours. This is in contrast to 17.4 percent of commuters who make one stop, 16.7 percent who make two stops, 8.8 percent reporting three stops, 5.5 percent reporting four stops, and 4.5 percent reporting five or more stops.

In summary, the data set offered a rich source of information and appropriate variation in dependent variables suitable for estimating a multi-dimensional choice model system with a mixture of dependent variable types. The model specification included a range of individual, household, and employment characteristics.

7.2 Modeling Methodology

This section presents a detailed description of the modeling methodology developed for estimating a multi-dimensional choice model system involving a mixture of dependent variable types. Figure 7.1a shows the various interdependencies that might exist in the choice continuum that this study intends to explore. The solid lines represent possible relationships within single time bands while the hollow lines represent relationships across temporal bands (scales). There can be joint decisions within a single temporal band as well as decisions that are interlinked across different temporal bands. Given that Castro *et al.* (2012) demonstrated how all traditional count models are special cases of generalized ordered response models, the methodology presented in this study is applicable for estimating choice systems that include count dependent variables as well. The model formulation accounts for correlated unobserved factors affecting multiple choice dimensions, and allows the estimation of all model parameters in a single step akin to classic full-information maximum likelihood approaches, thus ensuring the use of all information in parameter estimation leading to gains in efficiency. The remainder of this section presents the formulation.

7.2.1 Model Framework

Let there be G nominal variables for an individual, and let g be the index for the nominal variables ($g = 1, 2, 3, \ldots, G$)[11]. In the empirical context of the current research, G=3 (the

[11] A nominal variable can be an unordered multinomial response variable.

nominal variables are residential location, work location, and commute mode choice). Also, let I_g be the number of alternatives corresponding to the g^{th} nominal variable ($I_g \geq 3$) and let i_g be the corresponding index (i_g = 1, 2, 3, …, I_g). Note that I_g may vary across individuals, but index for individuals is suppressed at this time for ease of presentation. Also, it is possible that some nominal variables do not apply for some individuals, in which case G itself is a function of the individual q. However, the model is developed at the individual level, and so this notational nuance does not appear in the presentation here.

Consider the g^{th} nominal variable and assume that the individual under consideration chooses the alternative m_g. Also, assume the usual random utility structure for each alternative i_g.

$$U_{gi_g} = \boldsymbol{\beta}'_g \boldsymbol{x}_{gi_g} + \varepsilon_{gi_g}, \tag{1}$$

where $\boldsymbol{x}_{gi_g}$ is a (K_g×1)-column vector of exogenous attributes, $\boldsymbol{\beta}_g$ is a column vector of corresponding coefficients, and ε_{gi_g} is a normal error term. Let the variance-covariance matrix of the vertically stacked vector of errors $\varepsilon_g[=(\varepsilon_{g1}, \varepsilon_{g2}, ..., \varepsilon_{gI_g})']$ be $\boldsymbol{\Omega}_g$. As usual, appropriate scale and level normalization must be imposed on $\boldsymbol{\Omega}_g$ for identification. Under the utility maximization paradigm, $U_{gi_g} - U_{gm_g}$ must be less than zero for all $i_g \neq m_g$, since the individual chose alternative m_g. Let $u^*_{gi_gm_g} = U_{gi_g} - U_{gm_g} (i_g \neq m_g)$, and stack the latent utility differentials into a vector $\boldsymbol{u}^*_{\boldsymbol{g}} = \left[\left(u^*_{g1m_g}, u^*_{g2m_g}, ..., u^*_{gI_gm_g}\right)'; i_g \neq m_g\right]$. $\boldsymbol{u}^*_{\boldsymbol{g}}$ has a mean vector of $\boldsymbol{b_g} (\boldsymbol{\beta}'_{\boldsymbol{1}}\boldsymbol{z}_{\boldsymbol{g1m_g}}, \boldsymbol{\beta}'_{\boldsymbol{1}}\boldsymbol{z}_{\boldsymbol{g2m_g}}, ... \boldsymbol{\beta}'_{\boldsymbol{1}}\boldsymbol{z}_{\boldsymbol{gI_gm_g}})'$, where $\boldsymbol{z}_{\boldsymbol{gi_gm_g}} = \boldsymbol{x}_{\boldsymbol{gi_g}} - \boldsymbol{x}_{\boldsymbol{gm_g}}, i_g = 1, 2, ..., I_g; i_g \neq m_g$. To obtain the covariance matrix of $\boldsymbol{u}^*_{\boldsymbol{g}}$, define $\boldsymbol{M}_{\mathbf{g}}$ as an $(I_g - 1) \times I_g$ matrix that corresponds to an $(I_g - 1)$ identity matrix with an extra column of –1's added as the m_g^{th} column. Then, one may write:

$$\boldsymbol{u}^*_{\boldsymbol{g}} \sim N\left(\boldsymbol{b}_{\boldsymbol{g}}, \boldsymbol{\Sigma}^*_{\mathbf{g}}\right), \text{ where } \boldsymbol{\Sigma}^*_{\mathbf{g}} = \boldsymbol{M}_{\boldsymbol{g}}\boldsymbol{\Omega}_{\mathbf{g}}\boldsymbol{M}'_{\boldsymbol{g}}. \tag{2}$$

The discussion above focuses on a single nominal variable g. When there are G nominal variables, consider the stacked $\left[\sum_{g=1}^{G}(I_g - 1)\right] \times 1-$vector $\boldsymbol{u}^* = \left[\left(\boldsymbol{u}_1^{*'}, \boldsymbol{u}_2^{*'}, \ldots, \boldsymbol{u}_G^{*'}\right)'\right]$, each of whose element vectors is formed by differencing utilities of alternatives from the chosen alternative m_g for the g^{th} nominal variable. Next, one may write:

$\boldsymbol{u}^* \sim N(\boldsymbol{b}, \boldsymbol{\Sigma}^*)$, where $\boldsymbol{b} = \left(\boldsymbol{b}_1', \boldsymbol{b}_2', \ldots, \boldsymbol{b}_G'\right)'$ and $\boldsymbol{\Sigma}^*$ is a $\left[\sum_{g=1}^{G}(I_g - 1)\right] * \left[\sum_{g=1}^{G}(I_g - 1)\right]$ matrix as follows:

$$\boldsymbol{\Sigma}^* = \begin{bmatrix} \boldsymbol{\Sigma}_1^* & \boldsymbol{\Sigma}_{12}^* & . & . & . & \boldsymbol{\Sigma}_{1G}^* \\ \boldsymbol{\Sigma}_{21}^* & \boldsymbol{\Sigma}_2^* & . & . & . & \boldsymbol{\Sigma}_{2G}^* \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \boldsymbol{\Sigma}_{G1}^* & \boldsymbol{\Sigma}_{G2}^* & . & . & . & \boldsymbol{\Sigma}_G^* \end{bmatrix}$$

(3)

The off-diagonal elements in $\boldsymbol{\Sigma}^*$ capture the dependencies across the utility differentials of different nominal variables, the differential being taken with respect to the chosen alternative for each nominal variable.

Let there be L ordinal variables for an individual, and l be the index for the ordinal variables $(l = 1, 2, \ldots, L)$. In the empirical context of the current study, L=2 (the ordinal variables are vehicle ownership and number of stops in the commute). Also, let J_l be the number of outcome categories for the l^{th} ordinal variable $(J_l \geq 2)$ and let the corresponding index be j_l $(j_l = 1, 2, \ldots, J_l)$. Let y_l^* be the latent underlying variable whose horizontal partitioning leads to the observed choices for the l^{th} ordinal variable. Assume that the individual under consideration chooses the $n_l{}^{th}$ ordinal category. Then, in the usual ordered response formulation:

$$y_l^* = \boldsymbol{\delta}_l' \boldsymbol{w}_l + \xi_l,\ j_l = n_l \text{ if } \psi_{n_l - 1} < y_l^* < \psi_{n_l}, \tag{4}$$

where $\boldsymbol{w}_l$ is a vector of exogenous variables relevant to the l^{th} ordinal variable, $\boldsymbol{\delta}_l$ is a corresponding vector of coefficients to be estimated, the ψ terms represent thresholds, e_l is the index for the observed outcome for the ordinal variable $(j_l = 1, 2, \ldots, J_l)$, and ξ_l is the standard normal random error for the l^{th} ordinal variable. Stack the L latent variables y_l^* into an $(L \times 1)$ vector $\boldsymbol{y}^*$, and let $\boldsymbol{y}^* \sim N\left(\boldsymbol{f}, \boldsymbol{\Sigma}_{y^*}\right)$, where $\boldsymbol{f} = \left(= (\boldsymbol{\delta}_1' \boldsymbol{w}_l, \boldsymbol{\delta}_2' \boldsymbol{w}_2, \ldots\ldots \boldsymbol{\delta}_L' \boldsymbol{w}_L\right)$ and $\boldsymbol{\Sigma}_{y^*}$ is the covariance matrix of $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_L)$. Also, stack the lower thresholds $\psi_{n_l - 1}(l = 1, 2, \ldots, L)$ into an $(L \times 1)$ vector $\boldsymbol{\psi}_{low}$ and the upper thresholds $\psi_{n_l}(l = 1, 2, \ldots, L)$ into another vector $\boldsymbol{\psi}_{up}$.

Finally, let there be H continuous variables $(y_1, y_2, \ldots, y_H)$ with an associated index h $(h = 1, 2, \ldots, H)$. Let $y_h = \boldsymbol{\gamma}_h' \boldsymbol{s}_h + \eta_h$ in the usual linear regression fashion. Stacking the H continuous variables into a $(H \times 1)$-vector $\boldsymbol{y}$, one may write $y = N(\boldsymbol{c}, \boldsymbol{\Sigma}_y)$, where $\boldsymbol{c} = \left(\boldsymbol{\gamma}_1' \boldsymbol{s}_1, \boldsymbol{\gamma}_2' \boldsymbol{s}_2, \ldots\boldsymbol{\gamma}_H' \boldsymbol{s}_H\right)'$, and $\boldsymbol{\Sigma}_y$ is the covariance matrix of $\boldsymbol{\eta} = \left(\eta_1, \eta_2, \ldots \eta_H\right)$. The variance of $\tilde{\boldsymbol{y}} = \left(\boldsymbol{u}^*, \boldsymbol{y}^*, \boldsymbol{y}\right)$ can be written as:

$$\operatorname{Var}(\tilde{\boldsymbol{y}}) = \tilde{\boldsymbol{\Lambda}} = \begin{bmatrix} \boldsymbol{\Sigma}_{u^*} & \boldsymbol{\Sigma}_{u^* y^*} & \boldsymbol{\Sigma}_{u^* y} \\ \boldsymbol{\Sigma}_{u^* y^*} & \boldsymbol{\Sigma}_{y^*} & \boldsymbol{\Sigma}_{y^* y^*} \\ \boldsymbol{\Sigma}_{y^* y} & \boldsymbol{\Sigma}_{y^* y} & \boldsymbol{\Sigma}_{y} \end{bmatrix}, \tag{5}$$

where $\boldsymbol{\Sigma}_{u^* y^*}$ is a $\left[\sum_{g=1}^{G} (I_g - 1)\right] \times L$ matrix capturing covariance effects between the $\boldsymbol{u}^*$ vector and the $\boldsymbol{y}^*$ vector, $\boldsymbol{\Sigma}_{u^* y}$ is a $\left[\sum_{g=1}^{G} (I_g - 1)\right] \times H$ matrix capturing covariance effects between the $\boldsymbol{u}^*$ vector and the $\boldsymbol{y}$ vector, and $\boldsymbol{\Sigma}_{y^* y}$ is a $L \times H$ matrix capturing covariance effects between the $\boldsymbol{y}^*$ vector and the $\boldsymbol{y}$ vector. For ease in presentation, define $\tilde{\boldsymbol{u}} = \left(\boldsymbol{u}^{*\prime}, \boldsymbol{y}^{*\prime}\right)'$, $\tilde{g} = (b', f')'$, and $\boldsymbol{\Sigma}_{\tilde{u}} = \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{u}^*} & \boldsymbol{\Sigma}_{\boldsymbol{u}^* y^*} \\ \boldsymbol{\Sigma}_{\boldsymbol{u}^* y^*} & \boldsymbol{\Sigma}_{y^*} \end{bmatrix}$ and $\operatorname{Var}(\tilde{\boldsymbol{y}}) = \tilde{\boldsymbol{\Lambda}} = \begin{bmatrix} \boldsymbol{\Sigma}_{\tilde{u}} & \boldsymbol{\Sigma}_{\tilde{u} y} \\ \boldsymbol{\Sigma}_{\tilde{u} y} & \boldsymbol{\Sigma}_{y} \end{bmatrix}$.

Also, supplement the threshold vectors defined earlier as follows: $\tilde{\boldsymbol{\psi}}_{low} = \left[\left(-\infty_{\left[\sum_{g=1}^{G}(I_g-1)\right]} \right)', \boldsymbol{\psi}'_{low} \right]$, and $\tilde{\boldsymbol{\psi}}_{up} = \left[\left(\boldsymbol{0}_{\left[\sum_{g=1}^{G}(I_g-1)\right]} \right)', \boldsymbol{\psi}'_{up} \right]$, where $-\infty_M$ is a (M×1)-column vector of negative infinities, and $\boldsymbol{0}_M$ is another (M×1)-column vector of zeros. The conditional distribution of $\tilde{\boldsymbol{u}}$ given $\boldsymbol{y}$, is multivariate normal with mean $\tilde{\tilde{\boldsymbol{g}}} = \tilde{\boldsymbol{g}} + \boldsymbol{\Sigma}_{\tilde{u}y}\boldsymbol{\Sigma}_y^{-1}(\boldsymbol{y}-\boldsymbol{c})$ and variance $\tilde{\tilde{\boldsymbol{\Sigma}}}_{\tilde{\mathbf{u}}} = \boldsymbol{\Sigma}_{\tilde{\mathbf{u}}} - \boldsymbol{\Sigma}_{\tilde{\mathbf{u}}\mathbf{y}}\boldsymbol{\Sigma}_{\mathbf{y}}^{\mathbf{-1}}\boldsymbol{\Sigma}'_{\tilde{\mathbf{u}}\mathbf{y}}$.

Next, let $\boldsymbol{\theta}$ be the collection of parameters to be estimated: $\boldsymbol{\theta} = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, ..., \boldsymbol{\beta}_{\boldsymbol{G}};\ \boldsymbol{\delta}_1, \boldsymbol{\delta}_2, ..., \boldsymbol{\delta}_L;\ Vech(\boldsymbol{\Sigma}_{\tilde{u}});\ \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, ..., \boldsymbol{\gamma}_{\boldsymbol{H}};\ Vech(\boldsymbol{\Sigma}_y);\ Vech(\boldsymbol{\Sigma}_{\tilde{u}\mathbf{y}})]$, where $Vech(\mathbf{A})$ represents the vector of upper triangle elements of $\mathbf{A}$. Then the likelihood function for the individual may be written as:

$$L(\boldsymbol{\theta}) = \phi_H(\boldsymbol{y}-\boldsymbol{c} \mid \boldsymbol{\Sigma}_y) \times \Pr\left[\tilde{\boldsymbol{\psi}}_{low} \le \tilde{\boldsymbol{u}} \le \tilde{\boldsymbol{\psi}}_{up}\right], \tag{6}$$

$$= \phi_H(\boldsymbol{y}-\boldsymbol{c} \mid \boldsymbol{\Sigma}_y) \times \int_{D_{\tilde{u}}} \phi_{\tilde{G}+L}(\tilde{\boldsymbol{u}} \mid \tilde{\tilde{\boldsymbol{g}}}, \tilde{\tilde{\boldsymbol{\Sigma}}}_{\tilde{u}}) d\tilde{u},$$

where the integration domain $D_{\tilde{u}} = \{\tilde{\boldsymbol{u}} : \tilde{\boldsymbol{\psi}}_{low} \le \tilde{\boldsymbol{u}} \le \tilde{\boldsymbol{\psi}}_{up}\}$ is simply the multivariate region of the elements of the $\tilde{\boldsymbol{u}}$ vector determined by the vector of chosen alternatives in nominal variables and observed outcomes of ordinal variables, and $\phi_{\tilde{G}+L}(.)$ is the multivariate normal density function of dimension $\tilde{G}+L$, where $\tilde{G} = \left(\left[\sum_{g=1}^{G} (I_g - 1) \right] \right)$.

The above likelihood function involves the evaluation of a $\tilde{G}+L$-dimensional integral for each individual, which can be very computationally expensive if there are several nominal variables, or if each nominal variable can take a large number of values, or if there are several ordinal variables, or combinations of these. So, the Maximum Approximated Composite Marginal Likelihood (MACML) approach of Bhat (2011), in which the likelihood function only involves the computation of univariate and bivariate cumulative distributive functions, is used in this study.

7.2.2 The MACML Estimation Approach

Consider the following (pairwise) composite marginal likelihood function formed by taking the products (across the G nominal variables and L ordinal variables) of the joint pairwise probability of the chosen alternatives for an individual.

$$
\begin{aligned}
L_{CML}(\boldsymbol{\theta}) = \phi_H(\boldsymbol{y} - \boldsymbol{c} \mid \boldsymbol{\Sigma}_{\mathbf{y}}) \times & \left(\prod_{g=1}^{G-1} \prod_{g'=g+1}^{G} \Pr(d_{i_g} = m_g, d_{i_{g'}} = m_{g'}) \right) \\
& \times \left(\prod_{l=1}^{L-1} \prod_{l'=l+1}^{L} \Pr(j_l = n_l, j_{l'} = n_{l'}) \right) \\
& \times \left(\prod_{g=1}^{G} \prod_{l'=1}^{L} \Pr(d_{i_g} = m_g, j_l = n_l) \right).
\end{aligned} \tag{7}
$$

where d_{i_g} is an index for the individual's choice for the g^{th} nominal variable. The net result is that the pairwise likelihood function now only needs the evaluation of $\tilde{G}_{gg'}, \tilde{G}_{ll'}$, and $\tilde{G}_{gl}$ dimensional cumulative normal distribution functions (rather than the $\tilde{G}+L$-dimensional cumulative distribution function in the maximum likelihood function), where $\tilde{G}_{gg'} = I_g + I_{g'} - 2, \tilde{G}_{ll'} = 2,$ and $\tilde{G}_{gl} = I_g$. This leads to substantial computational efficiency. However, in cases where there are several alternatives for one or more nominal variables, the dimension $\tilde{G}_{gg'}$ and $\tilde{G}_{gl}$ can still be quite high. This is where the use of an analytic approximation of the multivariate normal cumulative distribution (MVNCD) function, as shown in Bhat (2011), is convenient. The resulting maximum approximated composite marginal likelihood (MACML) of Bhat (2011), which combines the CML approach with the analytic approximation for the MVNCD function evaluation, is solely based on bivariate and univariate cumulative normal computations. The MACML approach can be applied using a simple optimization approach for likelihood estimation. It also represents a conceptually simpler alternative to simulation techniques. Also, the MACML estimator $\hat{\boldsymbol{\theta}}_{MACML}$ is asymptotically normal distributed with mean $\boldsymbol{\theta}$ and covariance matrix given by the inverse of the Godambe's (1960) sandwich information matrix $G(\boldsymbol{\theta})$:

$$V_{MACML}(\boldsymbol{\theta}) = [G(\boldsymbol{\theta})]^{-1} = H(\boldsymbol{\theta})[J(\boldsymbol{\theta})]^{-1}[H(\boldsymbol{\theta})], \tag{8}$$

where $H(\boldsymbol{\theta})$ and $J(\boldsymbol{\theta})$ take the following form:

$$H(\boldsymbol{\theta}) = E\left[-\frac{\partial^2 \log L_{MACML}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right] \text{ and } J(\boldsymbol{\theta}) = E\left[\left(\frac{\partial \log L_{MACML}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial \log L_{MACML}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right)\right]$$

$H(\boldsymbol{\theta})$ and $J(\boldsymbol{\theta})$ can be estimated in a straightforward manner at the MACML estimate $\hat{\boldsymbol{\theta}}_{\boldsymbol{MACML}}$ as follows (introducing q as the index for individuals):

$$\hat{H}(\hat{\boldsymbol{\theta}}) = -\left[\sum_{q=1}^{Q} \frac{\partial^2 \log L_{MACML,q}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]_{\hat{\theta}_{CML}}, \text{ and}$$
$$\hat{J}(\hat{\boldsymbol{\theta}}) = \sum_{q=1}^{Q}\left[\left(\frac{\partial \log L_{MACML,q}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial \log L_{MACML,q}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right)\right]_{\hat{\theta}_{CML}}. \tag{9}$$

There are important identification and positive definiteness issues that must be taken into account during model estimation. These issues and the methods to deal with them are discussed in Chapter 6. In addition to the identification conditions discussed in that chapter, the scale of all ordinal variables must be normalized to one in the current model system to ensure identification.

7.3 Model Estimation Results

Model estimation results are described in this section. In the interest of brevity, only key findings and highlights of model estimation results are presented. In order to arrive at the final model specification, a number of model structures depicting alternative structural relationships among endogenous variables were estimated and examined with respect to statistical measures of fit. In the end, after extensive testing, plausibility checks, and goodness-of-fit assessment, the final model specification and set of structural relationships were identified. Figure 7.1b shows the relationships among dependent variables in the final model structure adopted in this study. It is found that residential location affects work location choice, both of which affect commute distance. All three long-term choice variables (residential location, work location, and commute distance) affect vehicle ownership. In turn, long term location choices and vehicle ownership influence commute mode choice which directly impacts trip chaining patterns (number of stops on commute). It should be noted that while there appears to be a sequential flow of relationships in this model specification, the model system itself is a joint simultaneous equations model that treats the set of dependent variables as a “bundle” with common unobserved effects affecting multiple choice dimensions. The jointness in the representation of the choice process is not lost due to the apparent sequential flow of relationships among variables.

In arriving at this specification, a variety of hypotheses were tested. For example, alternative structural relationships were examined with respect to long term choices – does residential location choice influence work location choice, or does work location affect residential location choice, or are residence and work location choices determined simultaneously, or does the (desired) commute distance affect residence and work locations? Similar hypotheses were constructed within and between temporal bands and alternative model structures were estimated. Finally, the model structure that is represented by Figure 7.1b was chosen as the best based on a variety of logic checks and goodness-of-fit measures.

7.3.1 Long Term Choice Model Components

Table 7.1 presents estimation results for long term choices. The residential location choice component of the model suggests that households with younger children have a greater propensity to locate in medium- to high-density neighborhoods, but households with older children shun the highest density neighborhoods, possibly in search of lower density suburban neighborhoods with good schools. Individuals with higher education levels favor residential locations in high density neighborhoods, suggesting that they may be interested in urban lifestyles that are more environmentally friendly. Lower income individuals tend to locate in high density neighborhoods while those seeking home ownership appear to do so in lowest density neighborhoods (likely to be in the suburbs). Immigrant households are more likely to favor higher density neighborhoods. The relative magnitude of the constants suggests that there is a baseline preference for low-to-medium density neighborhoods.

In terms of work location choice, it is found that there is a strong positive association between residential location density and work location density. It appears that people may be working in locations that are at least as dense as their residential neighborhoods, which is not surprising given that employment tends to locate such that workers can easily access jobs. Males are less likely to work in higher density locations than females. Individuals with higher education levels tend to find jobs in higher density areas (consistent with their residential location). Part time workers are more likely to work in high density areas, but as are self-employed individuals. It is possible that self-employed individuals seek high density areas where business opportunities abound. Immigrants are less likely to work in high density areas (in contrast to their residential location choice), but tend to favor higher density locations (similar to non-immigrant households) as they assimilate into the country over a period of time. Asians are less likely to work in higher density neighborhoods, while African Americans are more likely to do so.

The commute distance is similarly affected by a number of socio-economic variables. Males, full-time employees, and African Americans exhibit longer commutes, while lower income individuals and those with children have lower commuting distances. Those who own a home have longer commutes, presumably because they reside in distant suburbs to a greater degree. As residential location density or work location density increases, the commuting distance decreases; suggesting that there is an observed impact of density on commuting distance even after controlling for other factors and reflecting endogeneity through a simultaneous equations model system.

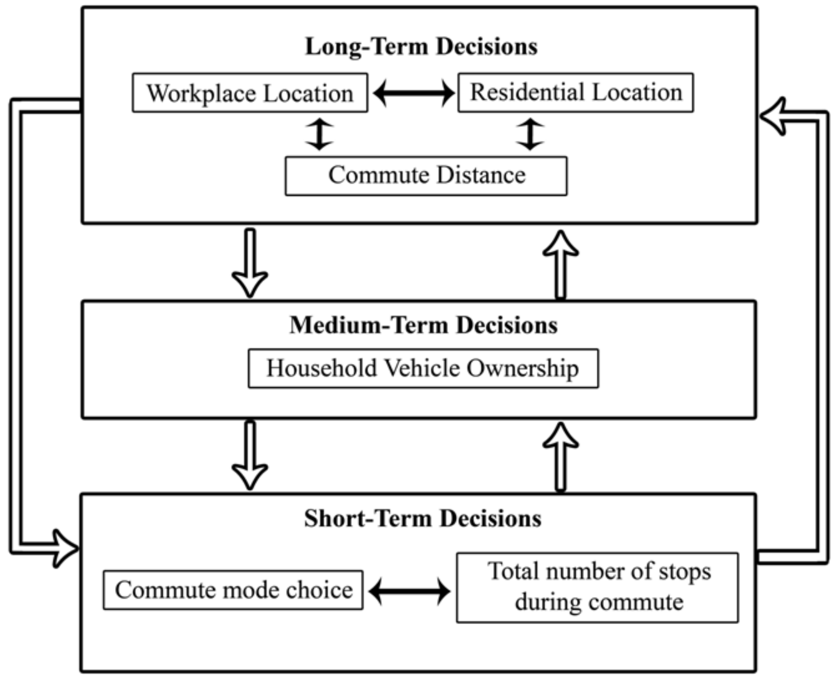

Figure 1a Possible interdependencies in the choice continuum.

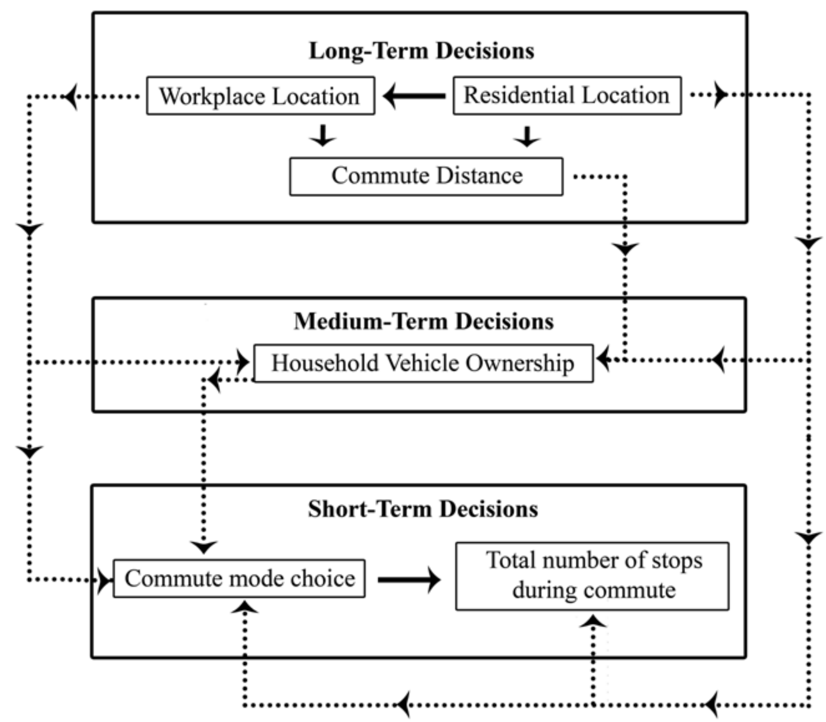

Figure 1b Relationships among endogenous variables in final model specification.

Figure 7.1 Interdependencies in the choice continuum

Table 7.1 Integrated Model Estimation Results – Long Term Choices

Variable Description	**Coef**	**t-stat**	**Coef**	**t-stat**	**Coef**	**t-stat**
Residential Location (Base Alternative: 0-499 housing units per square mile)	**500-1,999**		**2000-3,999**		**≥4,000**	
Constant	0.2413	2.10	0.2071	1.75	-0.0090	-0.05
Socio-economic Attributes						
Presence of children aged 6 to 10 years (Yes=1,No=0)			0.2427	2.14		
Presence of children aged 11-15 years (Yes=1,No=0)			0.2427	2.14	-0.4706	-3.96
Highest education attainment in household: College degree					0.4448	2.83
Highest education attainment in household: Post-doctoral degree					0.4807	3.05
Number of full time workers					0.1875	2.01
Number of self employed individuals			-0.1061	-1.77	-0.1061	-1.77
Number of workers with option to work from home	0.1889	2.70	0.1889	2.70	0.1889	2.70
Household income: Less than $20K (Yes=1 or No=0)			0.5132	2.71	0.5884	2.80
Housing tenure: Own house(Yes=1, No=0)	-0.1704	-1.41	-0.2501	-2.11	-0.8529	-7.12
Immigration status: Combination household			0.2153	2.67	0.2276	2.38
Immigration status: Immigrant household			0.1829	1.57	0.2378	1.87
Work Location (Base Alternative: 0-499 housing units per square mile)	**500-1,999**		**2000-3,999**		**>=4,000**	
Constant	-0.2174	-2.76	-0.7019	-4.33	-0.8493	-5.20
Socio-economic Attributes						
Gender (Male = 1, Female = 0)			-0.1199	-2.00	-0.1199	-2.00
Education attainment of the worker: College degree					0.1503	1.64
Education attainment of the worker: Post doctoral degree					0.1052	1.14
Full-time employment indicator (Yes = 1, No = 0)			-0.1270	-1.51	-0.1270	-1.51
Self employed (Yes=1, No=0)			0.5575	5.08	0.3336	2.54
Immigration status (Yes=1, No=0)			-0.2598	-2.70	-0.1614	-1.61
Immigration status: Number of years since entered the US	0.0047	1.51	0.0047	1.51	0.0047	1.51
Race of household respondent: African American					0.2687	1.35
Race of household respondent: Asian	-0.2033	-2.22	-0.2033	-2.22	-0.2033	-2.22
Residential Location						
500-1,999 housing units per square mile	0.2850	2.81	0.3915	3.50	0.3749	2.94
2,000-3,999 housing units per square mile	0.3451	3.32	0.6285	5.61	0.5331	4.11
≥4,000 housing units per square mile	0.3218	2.49	0.4793	3.27	1.1748	8.03
Natural Logarithm of Commute Distance (in miles)						
Constant	1.6760	13.28				
Socio-economic Attributes						
Gender (Male = 1, Female = 0)	0.2950	5.19				
Full-time employment indicator (Yes = 1, No = 0)	0.3970	5.69				
Self-employed (Yes=1,No=0)	-0.3960	-4.64				
Flexible work schedule (Yes=1, No=0)	0.1400	2.32				
Immigration status (Yes=1,No=0)	0.2490	3.13				
Race of household respondent: African American	0.4060	1.83				
Race of household respondent: Asian	-0.0860	-0.96				
Presence of children 0-15 years (Yes=1, No-0)	-0.0960	-1.52				
Household income: Less than $20K (Yes=1 or No=0)	-0.4590	-2.73				
Household income: $20K-$45K (Yes=1 or No=0)	-0.4690	-4.27				
Household income: $45K-$60K (Yes=1 or No=0)	-0.1830	-2.05				
Household income: $60K-$75K (Yes=1 or No=0)	-0.1730	-1.74				
Housing tenure: Own house(Yes=1, No=0)	0.2930	3.72				
Residential Location						
500-1,999 housing units per square mile	-0.1250	-1.64				
2,000-3,999 housing units per square mile	-0.2710	-3.30				
≥4,000 housing units per square mile	-0.5520	-5.55				
Work Location						
500-1,999 housing units per square mile	-0.1030	-1.40				
2,000-3,999 housing units per square mile	-0.0980	-1.33				
≥4,000 housing units per square mile	-0.0980	-1.33				

Table 7.2 Integrated Model Estimation Results – Medium Term Choice: Vehicle Ownership

Variable Description	Coef	t-stat
Thresholds		
Threshold 1 (1-2 vehicles)	-0.5866	-2.86
Threshold 2(2-3 vehicles)	0.9779	6.01
Threshold 3 (3-4 vehicles)	2.8345	17.35
Threshold 4 (4 or more vehicles)	3.8146	22.40
Socio-economic Attributes		
Number of adults in household	0.8614	20.76
Presence of children aged 11-15 years (Yes=1,No=0)	0.1481	1.71
Presence of senior adults aged 65 or over (Yes=1, No=0)	-0.2211	-2.27
Presence of an individual with prolonged medical condition (Yes=1, No=0)	-0.2293	-1.52
Highest education attainment in household: College degree	-0.2338	-2.87
Highest education attainment in household: Post-doctoral degree	-0.2997	-3.70
Number of full time workers	0.1524	2.74
Number of self employed individuals	0.1850	2.99
Number of individuals with more than one job	0.1322	1.46
Household income: Less than $20K (Yes=1 or No=0)	-0.7407	-5.13
Household income: $20K-$45K (Yes=1 or No=0)	-0.5459	-4.34
Household income: $45K-$60K (Yes=1 or No=0)	-0.3617	-3.28
Housing tenure: Own house(Yes=1, No=0)	0.7057	8.08
Residential Location		
500-1,999 housing units per square mile	-0.1078	-1.20
2,000-3,999 housing units per square mile	-0.1275	-1.39
≥4,000 housing units per square mile	-0.6695	-6.10
Work Location		
≥4,000 housing units per square mile	-0.2824	-3.16
Natural logarithm of commute distance (in miles)	0.0799	2.68

Table 7.3 Integrated Model Estimation Results – Short Term Choices

Variable Description	Coef	t-stat	Coef	t-stat	Coef	t-stat
Commute Mode (Base Alternative: Drive Alone)	**Shared Ride**		**Transit**		**Walk/Bike**	
Constant	-0.6794	-3.56	-2.8180	-11.51	-2.0918	-3.85
Socio-economic Attributes						
Age (in years)	-0.0143	-3.88			-0.0110	-1.52
Gender (Male = 1, Female = 0)	-0.0878	-1.05	0.1527	1.23	0.8706	4.08
Education attainment of the worker: Less than High school	0.3578	1.35	0.3578	1.35	0.3578	1.35
Self employed (Yes=1, No=0)	-0.1851	-1.27	-0.7354	-2.59	-0.9086	-2.33
Option to work from home (Yes=1, No=0)			0.2340	1.82	0.4062	1.52
Flexible work schedule (Yes=1, No=0)	0.2906	3.37	0.3087	2.43	0.6569	2.91
Immigration status (Yes=1, No=0)	0.2612	1.74	0.3838	2.18		
Immigration status: No. of years since entered US	-0.0053	-1.05	-0.0053	-1.05	-0.0053	-1.05
Race of household respondent: African American	-0.5177	-1.80	0.3201	1.25		
Residential Location						
500-1,999 housing units per square mile					0.8108	2.25
2,000-3,999 housing units per square mile			0.2191	1.57	0.8325	2.23
≥4,000 housing units per square mile	0.2193	1.93	0.9416	5.44	0.9486	2.43
Work Location						
500-1,999 housing units per square mile			0.2670	1.73		
2,000-3,999 housing units per square mile			0.6852	2.70	-0.6038	-2.49
≥4,000 housing units per square mile			0.6852	2.70	-0.6038	-2.49
Natural logarithm of Commute distance (in miles)			0.1555	2.17	-0.8513	-5.47
Vehicle Ownership						
Four or more vehicles	-0.1759	-3.89	-0.1759	-3.89	-0.1759	-3.89
Number of Commute Stops						
Thresholds						
Threshold 1 (1-2 stops)	0.2830	1.87				
Threshold 2(2-3 stops)	0.7738	5.04				
Threshold 3 (3-4 stops)	1.3366	8.57				
Threshold 4 (4 -5 stops)	1.7713	11.04				
Threshold 4 (5 or more vehicles)	2.2254	13.60				
Socio-economic Attributes						
Education attainment of the worker: College	0.1763	1.86				
Education attainment of the worker: Post-doctoral	0.1654	1.57				
Has more than one job (Yes=1,No=0)	0.3465	3.45				
Flexible work schedule (Yes=1, No=0)	0.3431	5.08				
Immigration status (Yes=1,No=0)	-0.1882	-2.23				
Race of household respondent: Caucasian	0.0946	1.18				
Presence of children 0-10 years (Yes=1, No-0)	0.1841	2.06				
Number of adults in household	-0.1894	-4.48				
Number of full time workers	0.1578	3.08				
Number of self employed individuals	0.2466	3.89				
Household income: Less than $20K (Yes=1 or No=0)	-0.2675	-1.46				
Household income: $20K-$45K (Yes=1 or No=0)	-0.2976	-2.17				
Residential Location						
≥4,000 housing units per square mile	0.1276	1.54				
Commute Mode						
Shared ride	0.6481	7.58				
Walk or bike	-0.5388	-2.48				

7.3.2 Medium Term Choice Model Component

The vehicle ownership model takes the form of an ordered response model. Higher levels of auto ownership are associated with a larger number of persons in the household. Thus, as number of adults, number of children, number of full time workers, number of self-employed individuals, and number of individuals with more than one job in the household (in which the sample respondent resides) increase, so does auto ownership. On the other hand, the presence of senior adults in the household or the prevalence of a medical condition has a negative impact on auto ownership presumably because these individuals have mobility limitations. As income levels fall, so do auto ownership levels as evidenced by the trend in negative coefficients associated with income dummy variables. Higher density residential location is associated with lower levels of auto ownership, presumably because these neighborhoods are better served by alternative modes and people who locate in such neighborhoods are not necessarily auto-oriented to begin with. Home ownership and longer commutes appear to contribute to higher levels of auto ownership. All of these indications are consistent with expectations.

7.3.3 Short Term Choice Model Component

Table 7.3 presents the model estimation results for the short-term choice components. There is negative baseline preference associated with the use of alternative modes of transport as evidenced by the negative constants. Older individuals are less likely to share a ride or bike/walk, possibly due to physical limitations. Males are less likely to share a ride, but more likely (than females) to use transit or bicycle and walk. Low education levels are associated with alternative mode use, possibly because these individuals are in low paying jobs, having lower income, and cannot afford to commute by car. Self-employed individuals are more likely to drive alone, possibly due to the flexibility that they need in seeking business opportunities. Those with a flexible work schedule are more likely to use alternative modes of transport. Immigrants are more likely to share a ride or use transit, but this effect dampens as the immigrants stay longer in the US and assimilate into the general population. Even after controlling for all other factors and endogeneity across choice dimensions, it is found that residential and workplace location density impact commute mode choice. Higher density location choices appear to contribute to greater levels of transit mode choice. Those working in high density tracts show a lower propensity to bicycle and walk, possibly because the areas are not conducive to non-motorized mode use (although conducive to transit use). High levels of vehicle ownership negatively impact alternative mode use.

The final dependent variable is the number of stops on the commute tours (an ordinal response variable). Consistent with expectations, higher levels of education, holding multiple jobs, and flexible work schedules are associated with higher levels of stop-making. Immigrants tend to make fewer stops, while Caucasians and individuals with children in the household tend to make more stops (due to serve-child trips). As the number of adults increases, stop making responsibilities are likely shared through household interactions, and stop-making propensity at the individual level drops. Similar task allocation effects are seen with respect to number of workers and number of self-employed individuals in the household. Lower income individuals tend to make fewer stops, possibly because their lower income does not afford them the opportunity to participate in as many discretionary activities. Those residing in the highest density neighborhoods engage in more stops, possibly because there are more destination opportunities that can be visited during the commute tour. In other words, higher residential density does not necessarily bring about inefficiencies in tour formation or activity engagement (where a person repeatedly returns home and starts a new tour to engage in new activities). Mode choice affects stop making behavior with those in shared ride mode likely to make more stops (to drop off and pick up passengers) and bicycle and walk commuters engaging in fewer stops, possibly in an effort to keep commuting distance and times manageable.

7.3.4 Self-Selection Effects and Model Assessment

Table 7.4 presents estimation results corresponding to the covariance matrix of utility differences, latent propensities, and continuous variables considered in this study. A number of interesting observations can be made. The significant parameter of 0.8009 in the block of covariances between modal utility differences, suggests that there are common unobserved factors affecting the choice of transit (relative to drive alone) and the choice of bicycle/walk (relative to drive alone). For example, people's attitudes about the environment and the desire to live a "green" lifestyle (which are unobserved effects) may be simultaneously (and positively) impacting preference for transit and bicycle/walk modes. There do not appear to be any significant endogeneity effects across residential and workplace location choices. The model estimation results revealed an observed impact of residential location choice on work location choice; there do not seem to be any common unobserved effects influencing these long term location choice decisions (at least in the context of this study).

Table 7.4 Covariance Matrix for the Integrated Model System

		Res (500-1,999)	Res (2,000-3,999)	Res (≥4,000)	Work (500-1,999)	Work (2,000-3,999)	Work (≥4,000)	Mode SR	Mode TR	Mode WB	# Veh	# Stops	Ln Comm Dist
Res	**(500-1,999)**	1.0											
Res	**(2,000-3,999)**	0.5	1.0										
Res	**(≥4,000)**	0.5	0.5	1.0									
Work	**(500-1,999)**	0.0	0.0	0.0	1.0								
Work	**(2,000-3,999)**	0.0	0.0	0.0	0.5	1.0							
Work	**(≥4,000)**	0.0	0.0	0.0	0.5	0.5	1.0						
Mode	**SR**	0.0	0.0	0.0	0.0	0.0	**0.0555 (1.03)**	1.0					
Mode	**TR**	0.0	0.0	0.0	0.0	0.0	**0.2883 (2.12)**	0.5	1.0				
Mode	**WB**	0.0	0.0	0.0	**-0.1507 (-1.1)**	0.0	0.0	0.5	**0.8009 (1.98)***	1.0			
# Veh		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	**-0.3317 (-2.09)**	1.0		
# Stops		**-0.0826 (-1.35)**	**0.0973 (1.72)**	0.0	0.0	0.0	**0.1004 (1.03)**	0.0	0.0	0.0	0.0	1.0	
Ln Comm Dist		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	**0.9804 (25.43)**

* T-statistic computed against 0.5 corresponding to the value in independent MNP model.

It appears that there are self-selection effects across work location choice and commute mode. The negative parameter of -0.1507 suggests that unobserved factors that contribute to a person choosing a low density area as work location are correlated with unobserved factors that make a person intrinsically less likely to walk or bicycle. These may be individuals who are more auto oriented by nature. Conversely, there are positive covariances (0.0555 and 0.2883) reflecting a positive disposition across the choice to work in high density areas and the choice of transit or shared ride as a commute mode. The unobserved factors that motivate an individual to seek a high density work place (desire for transit and pedestrian friendly options) are likely the very factors that contribute to higher level of transit and shared ride mode usage. Unobserved factors that contribute to an individual owning more vehicles (such as desiring an auto-oriented lifestyle) contribute negatively to the choice of bicycle and walk as a commute mode.

Similar self-selection effects are observed across residential location choice and number of stops, where it appears that the unobserved effects contributing to a choice of a high density residential location or work location positively impact stop-making behavior. This is plausible as an activity-seeking person who is an extrovert may choose residential and work locations that are high density (and provide such opportunities) and support their desire to engage in a variety of activities (stops) on the way to and from work.

The log-likelihood of the final model is -10508.1 and that of the model which ignores all potential correlations between the choices considered is -10520.4. The log-likelihood ratio test statistic of comparison between the two models is 24.54. This value is significantly greater than 15.51 which is the critical chi-squared value corresponding to 8 degrees of freedom at a 95 percent confidence level, thus demonstrating the superior statistical fit in the joint model.

7.4 Conclusions

This study constitutes a major attempt to build an integrated econometric model system of the choice continuum – connecting long term location choices, medium term vehicle ownership choice, and shorter term commute mode and commute tour stop-making behavior. The objective of this study is to advance the state of the art of econometric simultaneous equations modeling in the travel behavior and analysis domain with a view to better represent the bundling of choices that people may be exercising in their daily lives. People are not likely to make choices in isolation of one another or in a strictly sequential fashion, and there is widespread evidence that significant unobserved factors (that describe a person's attitudes, perceptions, experiences, and lifestyle preferences) simultaneously impact a multitude of choices.

The study presents an integrated econometric model system that ties together residential location choice, work location choice, commuting distance, vehicle ownership, commute mode choice, and number of stops made on commute tours. Thus, the model system includes a variety of dependent variable types commonly encountered in transport modeling contexts. The model system is estimated on a San Francisco Bay Area subsample of commuters drawn from the 2009 National Household Travel Survey data set in the United States. The study presents the model formulation and estimation procedures; recognizing that traditional estimation methods are computationally infeasible for the type of model system specified in this research, the study employs the maximum approximate composite marginal likelihood (MACML) estimation procedure together with a numerical approximation method to evaluate multi-dimensional integrals of the cumulative normal distribution function. The methodological breakthrough presented in this study offers the potential to bring integrated model systems of the nature estimated in this study closer to practical reality.

Model estimation results show that the choice dimensions considered in this study are inter-related, both through direct observed relationships and through correlations across unobserved factors (error terms) affecting multiple choice dimensions. The significant presence of self-selection effects (endogeneity) suggests that modeling the various choice processes in an independent sequence of models is not reflective of the true relationships that exist across these choice dimensions. The simultaneous equations model system estimated in this study offers superior fit in comparison to an independent set of models that ignore self-selection effects. Models that ignore self-selection effects are likely to provide erroneous forecasts of the impacts of land use density variables and policy strategies on activity-travel choices. The study findings suggest the following:

1. Residential location choice affects work location choice
2. Both residential and work location choices together impact commuting distance
3. Residential and work location choices, together with commuting distance, impact vehicle ownership
4. Both location choices, and vehicle ownership, affect commute mode choice
5. Commute mode choice and residential location affect number of stops on commute tours.

In addition to these observed relationships, the examination of error covariances shows that people with a propensity for non-auto oriented lifestyles (*i.e.*, greener lifestyles) tend to locate in higher density neighborhoods, adopt alternative modes of transport for their commute, and exhibit lower levels of automobile ownership. Clearly, attitudes and lifestyle preferences play an important role in shaping the multitude of choice dimensions considered and ignoring such self-selection effects can prove costly in policy forecasting and decision making processes. Future research efforts must be aimed at operationalizing integrated econometric model systems (such as that presented in this study) within activity-based travel forecasting models so that the types of endogeneity effects uncovered in this research can be better reflected in forecasts of travel demand under a wide range of modal and land use scenarios.

CHAPTER 8: Summary and Future Work

8.1 Summary of Work Undertaken

The objectives of this dissertation were three-fold. The first objective was to develop a comprehensive vehicle fleet composition and evolution microsimulation framework which can be embedded within activity-based travel demand models. The second objective was to develop state-of-the-art econometric models capable of (a) analyzing vehicle type, usage, and transaction decisions as a function of several alternative technology and policy scenarios, and (b) capturing the spatial dependencies in the vehicle purchase decisions of households. The third objective was to, develop integrated modeling frameworks of vehicle type/ownership, location, and daily activity-travel choices, and to capture the complex interactions among different choices that individuals make.

The comprehensive framework, the unique survey data used for the analysis, the copula-based modeling methodology, the vehicle spatial model development, the empirical application results, and the forecasting results discussed in Chapters 3, 4, and 5 address the first two objectives while the development of the integrated tour based model system and the integrated model system of residential location, work location, commute distance, vehicle ownership, and several commute tour choices described in Chapters 6 and 7 address the third objective of the dissertation.

8.1.1 Comprehensive Vehicle Composition and Evolution Framework

This dissertation has presented the design and formulation of a comprehensive vehicle fleet composition and evolution simulator that is capable of simulating household vehicle ownership and utilization decisions over time. The simulation framework consists of two main modules – one module that models the current (baseline) fleet composition and utilization for a household and another module that evolves the baseline fleet over time by considering the acquisition, replacement, and disposal processes that households may undertake as they turnover their fleet. The vehicle fleet simulator addresses several limitations of earlier studies. *First*, it accommodates all of the dimensions characterizing vehicle fleet/usage decisions. *Second*, it accommodates all of the dimensions of vehicle transactions (*i.e.*, fleet evolution) over time. *Third*, it allows multiple vehicle transaction in any given year. *Fourth*, it allows dependency between transaction decisions. *Finally*, it can be embedded easily within a larger travel microsimulation framework.

8.1.2 Unique Data & State-of-theArt Econometric Models

The residential survey component of the California Vehicle Survey data collected in 2008-2009 by the California Energy Commission (CEC) is used to estimate the model components in the vehicle fleet simulator. This data has three components -

1. Revealed choice (RC) component of the survey that collected detailed information on the current household vehicle fleet and usage
2. Stated intentions (SI) component to probe whether a household intended to replace an existing vehicle or acquire a net new additional vehicle in the fleet
3. Stated preference (SP) component including several vehicle types and fuel technology options not currently available in the market.

Data from these three survey components are pooled together to obtain a rich data set that can be used to model the full range of vehicle ownership and transactions decisions of households. This joint RC-SI-SP data makes the estimated models (and thus the entire vehicle fleet simulator) sensitive to technological evolutions and capable of examining the effects of a host of policy variables aimed at promoting vehicle mix/usage patterns that reduce GHG emissions and fuel consumption. The data also allowed using a very disaggregate dependent variable definition which is a combination of six vehicle body types (compact car, car, small cross utility vehicle, sport utility vehicle or SUV, van, and pick-up truck), seven fuel types (gasoline, flex fuel, plug-in hybrid, compressed natural gas (or CNG), diesel, hybrid electric, and fully electric), and five age categories (new, 1-2 years, 3-7 years, 8-12 years, and more than 12 years old).

Two important econometric issues that can have significant policy implications are addressed in the model components that constitute the vehicle fleet simulator. *First*, a copula-based joint modeling framework is applied to model vehicular choices and the associated vehicle mileage choices. The results show that the joint model performs substantially better than an independent set of model components that ignore common unobserved factors that impact both vehicle fleet composition and utilization. *Second,* a spatial lag multinomial probit model is developed to accommodate for the impact of the social/spatial interactions among households on the vehicle type choice decisions of the households. The resulting model is the first of its kind and has elasticity implications significantly different from a non-spatial model which ignores spatial dependencies. More importantly, the spatial vehicle type choice model fits perfectly into the microsimulation framework developed in this study.

8.1.3 Integrated Modeling

Tour-based model systems are increasingly being deployed to microsimulate daily activity-travel patterns of individuals. A host of tour attributes are modeled within these systems. However, a dimension that is often overlooked is the vehicle type choice. This dimension is of considerable importance in the energy consumption and emissions estimation arena. Another issue that arises is that most tour attributes in AB travel demand models are determined independently or sequentially with loose coupling across the models, thus ignoring important endogeneity effects that may exist across multiple tour dimensions. This dissertation considers four key dimensions of tours – tour complexity, passenger accompaniment, vehicle type choice, and tour length – with a view to developing a joint simultaneous equations model system of tour choices while accounting for the presence of correlated unobserved attributes affecting multiple dimensions through appropriate error covariance structures.

In the same spirit, day-to-day activity travel choices are made in combination with other decisions such as residential location, work location, and vehicle ownership. Households and individuals also may select a specific residential location, work location, and vehicle ownership portfolio based on their desired activity-travel characteristics. In addition to this inter-relationship between the location and vehicle ownership decisions and activity-travel choices, the individual decisions of residential location, work location, and vehicle ownership themselves are likely to be inter-linked. Similarly, many studies in the literature indicate that households and individuals may locate themselves into neighborhoods based on their desire to own vehicles (*e.g.*, auto-disinclination), and travel preferences (*e.g.*, preference for walking). As it is conceivable that several choices are bundled together and made jointly by individuals, there is a need to simultaneously model a multitude of choices in an integrated framework. The specification and estimation of such simultaneous equations model systems has remained a challenge and prevented progress in this domain. This dissertation addresses this challenge by offering an econometric model system that simultaneously considers six different choice dimensions in a unifying framework. The six dimensions include residential location choice, work location choice, auto ownership, commuting distance, commute mode, and number of stops on commute tours. The dimensions modeled in this study cover disparate temporal scales from the long term location choices to the short term activity-travel choices, and include a variety of dependent variable types. Estimation results show a substantial presence of correlated unobserved effects (self-selection) across choice dimensions, underscoring the value offered by simultaneous equations model systems in the travel modeling field.

8.2 Directions for Future Research

There are several key issues that can further improve our understanding of the intricate relationships among different choices that individuals make. Some of these research ideas are discussed below:

8.2.1 Residential Self-Selection

Many studies in the literature (for example, Bhat and Guo, 2007) indicate that households and individuals may locate themselves into neighborhoods based on their desire to own certain type of vehicles (this desire may be a function of socio-demographics (*e.g.*, income), attitudes (*e.g.*, large auto-disinclination), and travel preferences (*e.g.*, preference for walking)). Although such residential self-selection effects are considered in the integrated residential location, work location, vehicle ownership, and commute tour characteristics model described in Chapter 7, they are not considered in the vehicle type choice models primarily due to data limitations. Future data collection efforts and modeling methods exploring the residential self-selection efforts will help uncover the true impact of the built environment on vehicle type choice decisions.

The vehicle spatial model developed in this dissertation also does not accommodate for residential self-selection effects. These residential self-selection effects, when ignored, can manifest themselves as spurious social/spatial interaction effects (*i.e.,* those caused by dyadic interactions of individuals located in close social or spatial proximity). Households may have different intrinsic preferences for specific types of vehicles due to unobserved lifestyle preferences and unobserved location factors, and may also have different sensitivities to exogenous variables such as gasoline costs, vehicle purchase prices, and residential built environment attributes (land-use mix, transit availability and accessibility *etc.*). This would then translate to a household-specific random coefficients formulation, leading also to a stationary across-time correlation for the same household in the case of pseudo-panel data with synthetic vehicle purchase choice occasions. Ignoring the presence of such unobserved heterogeneity will, in general, lead to biased and inconsistent parameters on all model parameters, including the spatial lag effect. In addition, because of the spatial nature of household locations, some earlier studies have suggested that these unobserved heterogeneity effects should be correlated over households based on the spatial (or social) proximity of the residential locations, which is then referred to as spatial drift (see Bradlow *et al.*, 2005 for a discussion). For example, Bernardo *et al.* (2012) recently developed a multinomial probit spatial drift model to analyze mode choice decisions. Future research can benefit significantly by examining for the presence of spatial drift effects in the vehicle type choice decisions of households.

8.2.2 Revealed Preference Data

Although the share of alternative fuel vehicles is increasing, a majority of the vehicles on the road today are still fueled by conventional fuels such as gasoline and diesel. This is one of the main reasons this study used the stated preference (SP) data to understand how their preferences for vehicles with small market penetration today would change under future market conditions where more alternate fuel options become available. Although we control for the possibility that the choice process exhibited in the revealed preference (RP) data is different from that exhibited in the SP data, the information regarding alternate fuel vehicle (AFV) purchasing behavior comes predominantly from the SP data. Like any other SP data based model, there is possibility that the vehicle type choice model developed in this study does not capture actual market behavior. Joint RP-SP data models of vehicle type choice with a significant RP component on AFV purchase behavior can substantially enhance our understanding of the actual vehicle purchasing behavior of households. Given the increasing share of AFVs, future data collection efforts must focus on capturing changing vehicle and fuel preferences.

8.2.3 Well-to-Wheel Emissions

In the vehicle fleet forecasts described in Chapter 4 of the dissertation, only the emissions associated with vehicle operation are considered. For instance, although fully electric vehicles produce zero emissions during operation, there are emissions associated with car manufacturing and electricity generation. A more comprehensive well-to-wheel analysis is needed to provide a holistic understanding of the associated costs and benefits (Thomas, 2012).

8.2.4 Intra-household Interactions

The decisions regarding residential location, work location, work schedules, and work-related travel activities are not made in isolation. For instance, Khan *et al.* (2012) found that the work arrangement decisions (including employed or not, work full-time or part-time, be self-employed or not, hold more than one job or not, work at home or not) of all individuals in a household are not independent due to the presence of correlated unobserved factors both within and between individuals in the household. The multivariate modeling framework developed in this study can be extended to explore these complex interactions among household members and its impact on vehicle ownership and activity-travel patterns.

Given that the percentage of households with dual earning members is increasing, it is only reasonable to assume that there are multiple people in the household who influence decisions such as residential location and vehicle ownership/ type choices. The final outcome can be viewed as an outcome of a cooperative bargaining process in the household with different individuals having varying travel needs and influences. For instance, Zhang *et al*., 2005 and Kato and Matsumoto, 2009 use group decision-making approach to analyze the time-use decisions of all individuals in a household. It would be interesting to see the differences in the model implications between the single household-level utility based approach (which is the one used in this dissertation) and a multiple individual-level utility based approach where 1) every member derives different utility from a vehicle type alternative, and 2) each member influences each vehicle type alternative's utility of every other household member to a different degree.

References

Abay, K.A., R. Paleti, and C.R. Bhat (2012) The joint analysis of injury severity of drivers in two-vehicle crashes accommodating seat belt use endogeneity. Technical paper, The University of Texas at Austin.

Abraham, J. and J. Hunt (1997) Specification and estimation of nested logit model of home, workplaces, and commuter mode choices by multiple-worker households. *Transportation Research Record: Journal of the Transportation Research Board*, 1606, 17-24.

ACS (2011) Marital Events of Americans: 2009, American Community Survey Reports. Available at: http://www.census.gov/prod/2011pubs/acs-13.pdf

Adjemian, M.K., C.Y.C Lin, and J. Williams (2010) Estimating spatial interdependence in automobile type choice with survey data. *Transportation Research Part A*, 44(9), 661-675.

Anderson, P.H., J.A. Mathews, and M. Rask (2009) Integrating private transport into renewable energy policy: The strategy of creating intelligent recharging grids for electric vehicles, *Energy Policy*, 37, 2481-2486.

Anselin, L. (2003) Spatial externalities, spatial multipliers and spatial econometrics. *International Regional Science Review*, 26(2), 153-166.

Anselin, L. (2006) Spatial econometrics. In: Mills T., Patterson K. (eds), *Palgrave Handbook of Econometrics: Volume 1, Econometric Theory*, Palgrave Macmillan Basingstoke, 901-969.

Anselin, L. (2010) Thirty years of spatial econometrics. *Papers in Regional Science*, 89(1), 3-25.

Bagley, M.N. and P.L. Mokhtarian (2002) The impact of residential neighborhood type on travel behavior: A Structural Equations Modeling Approach. *The Annals of Regional Science*, 36(2), 279-297.

Bernardo, C., R. Paleti, and C.R. Bhat (2012) A new spatial (social) interaction discrete choice model accommodating self-selection in group formation. Technical Paper. The University of Texas at Austin.

Beron, K.J., and W.P.M. Vijverberg (2004) Probit in a spatial context: A Monte Carlo Analysis. In L.Anselin, R.J.G.M. Florax, & S.J. Rey, eds., Advances in Spatial Econometrics: Methodology, Tools and Applications. Berlin: Springer-Verlag.

Beron, K.J., J.C. Murdoch, and W.P.M. Vijverberg (2003) Why cooperate? Public goods, economic power, and the Montreal protocol. *Review of Economics and Statistics* 85(2), 286-97.

BFIN (2012) Oak Ridge National Laboratory. Retrieved from: http://bioenergy.ornl.gov/papers/misc/energy_conv.html

Bhat, C.R. (2011) The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models. *Transportation Research Part B*, 45(7), 923-939.

Bhat, C.R., I.N. Sener, and N. Eluru (2010) A flexible spatially dependent discrete choice model: formulation and application to teenagers' weekday recreational activity participation. *Transportation Research Part B*, 44(8-9), 903-921.

Bhat, C.R. and N. Eluru (2010) Recent advances in discrete and discrete-continuous modeling systems. In S. Nocera (ed) *Feasibility Decisions in Transportation Engineering*, McGraw-Hill, 111-114.

Bhat C.R., S. Sen, N. Eluru (2009) The impact of demographics, built environment attributes, vehicle characteristics, and gasoline prices on household vehicle holdings and use. *Transportation Research Part B* 43(1): 1-18.

Bhat, C.R. and N. Eluru (2009) A copula-based approach to accommodate residential self-selection effects in travel behavior modeling. *Transportation Research Part B*, 43(7), 749-765.

Bhat, C.R., N. Eluru, and R.B. Copperman (2008) Flexible model structures for discrete choice analysis. In D.A. Hensher and K.J. Button (eds) *Handbook of Transport Modelling*, 2nd edition, Elsevier Science, 75-104.

Bhat, C.R., and J.Y. Guo (2007) A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels. *Transportation Research Part B*, 41(5), 506-526.

Bhat, C.R. and S. Sen (2006) Household vehicle type holdings and usage: an application of the multiple discrete-continuous extreme value (mdcev) model. *Transportation Research Part B*, 40(1), 35-53.

Bhat, C.R. and S. Srinivasan (2005) A multidimensional mixed ordered-response model for analyzing weekend activity participation. *Transportation Research Part B*, 39(3), 255-278.

Bhat, C.R., J.Y. Guo, S. Srinivasan, and A. Sivakumar (2004) Comprehensive econometric microsimulator for daily activity-travel patterns. *Transportation Research Record: Journal of the Transportation Research Board*, 1894, 57-66.

Bhat, C.R. and H. Zhao (2002) The spatial analysis of activity stop generation. *Transportation Research Part B*, 36(6), 557-575.

Bhat, C.R. (1998) A model of post home-arrival activity participation behavior. *Transportation Research Part B*, 32(6), 387-400.

Borjesson, M. (2008) Joint RP–SP data in a mixed logit analysis of trip timing decisions. *Transportation Research Part E*, 44(6), 1025-1038.

Bowman, J.L. and M.E. Ben-Akiva (2001) Activity-based disaggregate travel demand model system with activity schedules. *Transportation Research Part A*, 35(1), 1-28.

Bradlow, E.T., B. Bronnenberg, G. J. Russell, N. Arora, D. R. Bell, S. D. Duvvuri, F. T. Hofstede, C. Sismeiro, R. Thomadsen, & S. Yang (2005) Spatial models in marketing. *Marketing Letters*, 16 (3), 267-278.

Brown, B. (1986) Modal choice, location demand, and income. *Journal of Urban Economics*, 20(2), 128-139.

Brownstone, D., D. S. Bunch, and K. Train (2000) Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles. *Transportation Research Part B*, 34(5), 315-338.

Brownstone, D. and T. F. Golob (2009) The impact of residential density on vehicle usage and energy consumption. *Journal of Urban Economics*, 65(1), 91-98.

Bunch, D. S. (2000) Automobile demand and type choice. In D. A. Hensher and K. J. Button (Eds) *Handbook of Transport Modelling*, pp. 463-479, Pergamon, Oxford, 2000.

Cao, X., P.L. Mokhtarian, and S.L. Handy (2008) Examining the impacts of residential self-selection on travel behaviour: a focus on empirical findings. *Transport Reviews*, 29(3), 359-395.

Cao, X., P.L. Mokhtarian, S.L. Handy (2009) The relationship between the built environment and nonwork travel: a case study of Northern California. *Transportation Research Part A*, 43(5), 548-559.

Castro, M., R. Paleti, and C.R. Bhat, C.R. (2012) A latent variable representation of count data models to accommodate spatial and temporal dependence: application to predicting crash frequency at intersections. *Transportation Research Part B*, 46(1), 253-272.

CDPH (2010) Death Statistical Data Tables, Califnornia Department of Public Health. Available at: http://www.cdph.ca.gov/data/statistics/Pages/DeathStatisticalDataTables.aspx

Chen, Y., S. Ravulaparthy, K. Deutsch, P. Dalal, S.Y. Yoon, T. Lei, K.G. Goulias, R.M. Pendyala, C.R. Bhat, and H.H. Hu (2011) Development of indicators of opportunity-based accessibility. *Transportation Research Record: Journal of the Transportation Research Board*, 2255, 58-68.

Choo, S., and P.L. Mokhtarian (2004) What type of vehicle do people drive? The role of attitude and lifestyle in influencing vehicle type choice. *Transportation Research Part A*, 38(3), 201-222.

Chung, J. and V.R. Rao (2003) A general choice model for bundles with multiple-category products: application to market segmentation and optimal pricing for bundles. *Journal of Marketing Research*, 40(2), 115-130.

Dargay J.M., and P.C. Vythoulkas (1999) Estimation of a dynamic car model: a pseudo-panel approach. *Journal of Transport Economics and Policy*, 33(3), 287-302.

de Haan, P., M. G. Mueller, and R. W. Scholz (2009) How much do incentives affect car purchase? Agent-based microsimulation of consumer choice of new cars, part ii: Forecasting effects of feebates based on energy-efficiency. *Energy Policy*, 37(3), 1083-1094.

de Jong, G. and R. Kitamura (1992) A review of household dynamic vehicle ownership models: holdings models versus transactions models. *Proceedings of Seminar E, 20th PTRC Summer Annual Meeting,* PTRC Education and Research Services Ltd., London, 141-152.

DeSalvo, J.S., and M. Huq (1996) Income, residential location, and mode choice. *Journal of Urban Economics*, 40(1), 84-99.

DeSalvo, J.S., and M. Huq (2005) Mode choice, commuting cost, and urban household behavior. *Journal of Regional Science*, 45(3), 493-517.

DOF (2011) Historical and projected state and county births, 1980-2020, with actual and projected fertility rates by mother's age and race/ethnicity, 2000-2020. Available at: http://www.dof.ca.gov/research/demographic/reports/projections/births/

DOF (2012) Interim population projections for California and its counties 2010-2050. Available at: http://www.dof.ca.gov/research/demographic/reports/projections/interim/view.php

Dugundji, E.R., and J.L. Walker (2005) Discrete choice with social and spatial network interdependencies: an empirical example using mixed generalized extreme value models with field and panel effects. *Transportation Research Record: Journal of the Transportation Research Board*, 1921, 70-78.

Elhorst, J.P. (2010) Spatial panel data models. In: Fischer, M.M., Getis, A. (eds.), *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*, Springer, Berlin, 377- 408.

Eluru, N., Bhat, C.R., Pendyala, R.M., & Konduri, K.C. (2010) A joint flexible econometric model system of household residential location and vehicle fleet composition/usage choices. *Transportation*, 37(4), 603-626.

Eluru, N., Sener, I.N., Bhat, C.R., Pendyala, R.M., and K.W. Axhausen (2009) Understanding residential mobility. *Transportation Research Record: Journal of the Transportation Research Board*, 2133, 64-74.

Eluru, N., A.R. Pinjari, J.Y. Guo, I.N. Sener, S. Srinivasan, R.B. Copperman, and C.R. Bhat (2008) Population updating system structures and models embedded in the comprehensive econometric microsimulator for urban systems, *Transportation Research Record*, 2076, 171-182.

EPA (2011) *Inventory of U.S. Greenhouse Gas Emissions and Sinks 1990-2009*. Environmental Protection Agency and U.S. Department of Energy, 2011.

EPA (2010) Sources of greenhouse gas emissions: transportation sector emissions. United States Environmental Protection Agency, Washington, D.C. Available at: http://epa.gov/climatechange/ghgemissions/sources/transportation.html

EPA (2007) Climate Change: Human-related sources and sinks of carbon dioxide. Environmental Protection Agency and U.S. Department of Energy, 2007.

EPA (2006) Greenhouse gas emissions from the u.s. transportation sector 1990–2003, *Office of Transportation and Air Quality* (6401A) EPA 420 R 06 003 March 2006. www.epa.gov/otaq/climate.htm.

EPA (2005) Emissions facts: Greenhouse gas emissions from a typical passenger vehicle, *Office of Transportation and Air Quality* (6401A) EPA420-F-05-004 February 2005. http://www.epa.gov/otaq/climate/420f05004.pdf

Fang, A. (2008) Discrete-continuous model of households' vehicle choice and usage, with an application to the effects of residential density. *Transportation Research Part B*, 42(9), 736-758.

Franzese, R.J., Hays, J.C., and L.M. Schaffer (2010) Spatial, temporal, and spatiotemporal autoregressive probit models of binary outcomes: estimation, interpretation, and presentation. APSA 2010 Annual Meeting Paper. Available at: http://ssrn.com/abstract=1643867

Godambe, V. (1960) An optimum property of regular maximum likelihood estimation. *Annals of Mathematical Statistics*, 31, 1208-1211.

Golob, T.F. (2003) Structural equation modeling for travel behavior research. *Transportation Research Part B*, 37(1), 1-25.

Golob, T.F. (2000) A simultaneous model of household activity participation and trip chain generation. *Transportation Research Part B*, 34(5), 355-376.

Hamed, M.M. and F.L. Mannering (1993) Modeling traveler's postwork activity involvement: Toward a new methodology. *Transportation Science*, 27(4), pp. 381-394.

Hensher, D. A. and V. Le Plastrier (1985) Towards a dynamic discrete-choice model of household automobile fleet size and composition. *Transportation Research Part B*, 19(6), 481-495.

Kato, H. and M. Matsumoto (2009) Intra-household interaction in a nuclear family: a utility-maximizing approach. *Transportation Research Part B*, 43 (2), 191-203.

Khan, M., R. Paleti, C.R. Bhat, and R.M. Pendyala (2012) A Joint Household Level Analysis of Work Arrangement Choices of Individuals. Forthcoming, *Transportation Research Record*.

Kitamura, R., T. Yamamoto, S. Fujii, and S. Sampath (1996) A Discrete-continuous analysis of time allocation to two types of discretionary activities which accounts for unobserved heterogeneity. In J.B. Lesort (ed) *Transportation and Traffic Theory*, Pergamon Press, Oxford, UK, 431-453.

Konduri, K.C., X. Ye, B. Sana, and R.M. Pendyala (2011) A joint tour-based model of vehicle type choice and tour length. *Transportation Research Record*: *Journal of the Transportation Research Board*, 2255, 28-37.

Krizek, J.K. (2003) Neighborhood services, trip purpose, and tour-based travel. Transportation 30(4), 387-410.

Lave, C.A. and K. Train (1979) A disaggregate model of auto-type choice. *Transportation Research Part A*, 13(1), 1-9.

Lee, L-F. and J. Yu (2010) Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics* 154(2), pp. 165-185.

Lee L. F. (1983) Generalized econometric models with selectivity. *Econometrica*, 51(2), 507-512.

Lerman, S.R. (1976) Location, housing, automobile ownership and mode to work: a joint choice model. *Transportation Research Record*: *Journal of the Transportation Research Board*, 610, 6-11.

LeRoy, S.F. and J. Sonstelie. (1983) Paradise lost and regained: Transportation innovation, income, and residential location. *Journal of Urban Economics*, 13(1), 67-89.

LeSage, J.P. and R.K. Pace (2009) Introduction to Spatial Econometrics. Chapman & all/CRC, Taylor & Francis Group, Boca Raton.

Mannering, F.L. and D.A. Hensher (2007) Discrete/Continuous econometric models and their application to transport analysis. *Transport Reviews*, 7(3), 227-244.

Mannering, F., C. Winston, and W. Starkey (2002) An exploratory analysis of automobile leasing by US households. *Journal of Urban Economics*, 52(1), 154-176.

Mannering, F.L. and C. Winston (1985) A Dynamic empirical analysis of household vehicle ownership and utilization. *The RAND Journal of Economics*, 16(2), 215-236.

Mohammadian, A. and E.J. Miller (2003) Dynamic modeling of household automobile transactions. *Transportation Research Record: Journal of the Transportation Research Board*, 1831, 98-105.

Mohammadian, A. and E.J. Miller (2003) Empirical investigation of household vehicle type choice decisions. *Transportation Research Record: Journal of the Transportation Research Board*, 1854, 99-106.

Mohammadian A., M. Haider and P.S. Kanaroglou (2005) Incorporating spatial dependencies in random parameter discrete choice models. Presented at the *84th Annual Meeting of the Transportation Research Board*, Washington D.C.

Mokhtarian, P.L. and X. Cao (2008) Examining the Impacts of Residential Self-Selection on Travel Behavior: A Focus on Methodologies. *Transportation Research Part B*, 42(3), 204-228.

Mueller, M. G. and P. de Haan (2009) How much do incentives affect car purchase? Agent-based microsimulation of consumer choice of new cars, Part I: Model structure, simulation of bounded rationality, and model validation. *Energy Policy*, 37(3), 1072-1082.

Musti, S. and K.M. Kockelman. (2011) Evolution of the household vehicle fleet: Anticipating fleet composition, PHEV adoption and GHG emissions in Austin, Texas. *Transportation Research Part A*, 45(8), 707-720.

NAS (2008) *What You Need to Know About Energy?* National Academy of Sciences, The National Academies.

National Conference of State Legislatures (2011) State electric vehicle incentives. Available at: http://www.ncsl.org/issues-research/energyhome/state-electric-vehicle-incentives-state-chart.aspx

NHTSA (2012) Fact sheet: CAFE Fuel Economy. Available at: http://www.nhtsa.gov/staticfiles/rulemaking/pdf/cafe/2017-25_CAFE_NPRM_Factsheet.pdf

Noland, R.B. and W.A. Cowart (2000) Analysis of metropolitan highway capacity and the growth in vehicle miles of travel. *Transportation*, 27(4), 363-390.

Páez, A. and D. Scott (2004) Spatial statistics for urban analysis: a review of techniques with examples. *GeoJournal*, 61(1), 53-67.

Paleti, R., R.B. Copperman, and C.R. Bhat (2011) An empirical analysis of children's after school out-of-home activity-location engagement patterns and time allocation. *Transportation*, 38(2), 273-304.

Paleti, R., C.R. Bhat, and R.M. Pendyala, (2012) An integrated model of residential location, work location, vehicle ownership, and commute tour characteristics. Technical Paper, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin.

Paleti, R., C.R. Bhat, R.M. Pendyala, and K.G. Goulias, (2012) The Modeling of Household Vehicle Type Choice Accommodating Spatial Dependence Effects. Forthcoming, *Transportation Research Record.*

Paleti, R., N. Eluru, C.R. Bhat, R.M. Pendyala, T.J. Adler, and K.G. Goulias (2011) Design of comprehensive microsimulator of household vehicle fleet composition, utilization, and evolution. *Transportation Research Record*, 2254, 44-57.

Paleti, R., R.M. Pendyala, C.R. Bhat, K.C. Konduri, (2011) A joint tour-based model of tour complexity, passenger accompaniment, vehicle type choice, and tour length. Technical Paper, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin.

Pendyala, R.M., C.R. Bhat, K.G. Goulias, R. Paleti, K.C. Konduri, R. Sidharthan, H. Hu, G. Huan, and K.P. Christian (2012) The application of a socio-economic model system for activity based modeling: experience from southern California. Forthcoming, *Transportation Research Record*.

Pendyala, R.M. and C.R. Bhat (2004) An exploration of the relationship between timing and duration of maintenance activities. *Transportation*, 31(4), 429-456.

Phaneuf, D.J. and R.B. Palmquist (2003) Estimating spatially and temporally explicit land conversion models using discrete duration. Presented at the *AERE 2003 Summer Workshop* on Spatial Theory, Modeling, and Econometrics in Environmental and Resource Economics, June 15-17, Madison, WI.

Pinjari, A.R., R.M. Pendyala, C.R. Bhat, and P.A. Waddell (2011) Modeling the choice continuum: an integrated model of residential location, auto ownership, bicycle ownership, and commute tour mode choice decisions. *Transportation*, 38(6), 933-958.

Pitkin, J. and D. Myers (2012) Generational projections of the California population by nativity and year of immigrant arrival. Produced by the Population Dynamics Research Group, Sol Price School of Public Policy, University of Southern California. Text and supporting materials are published at: http://www.usc.edu/schools/price/research/popdynamics

Pew Research Centre (PRC) (2012) The Boomerang Generation feeling ok about living with mom and dad. Available at: http://www.pewsocialtrends.org/2012/03/15/the-boomerang-generation/2/#who-are-the-boomerang-kids?src=prc-number

Prillwitz, J., S. Harms, and M. Lanzendorf (2006) Impact of life-course events on car ownership. In *Transportation Research Record: Journal of the Transportation Research Board*, 1985, Transportation Research Board of the National Academies, Washington, D.C., 71-77.

Roorda, M. J., J. A. Carrasco, and E. J. Miller (2008) An integrated model of vehicle transactions, activity scheduling and mode choice. *Transportation Research Part B*, 43(2), 217-229.

Shiftan, Y. and J. Suhrbier (2002) The analysis of travel and emission impacts of travel demand management strategies using activity-based models. *Transportation*, 29(2), 145-168.

Smirnov, O.A. (2010) Spatial econometrics approach to integration of behavioral biases in travel demand analysis. *Transportation Research Record* 2157, 1-10.

Smith, T. and J.P. LeSage (2004) A Bayesian probit model with spatial dependencies. In LeSage J.P., Pace, R.K. (eds), *Advances in Econometrics, Volume 18: Spatial and Spatiotemporal Econometrics*, 127-160, Elsevier Ltd, Oxford.

Thomas, S. (2012) How green are electric vehicles? *International Journal of Hydrogen Energy*, 37, 6053-6062.

Train, K. (1980) The potential market for non-gasoline-powered automobiles. *Transportation Research Part A*, 14(5-6), 405-414.

Trivedi, P. K. and D. M. Zimmer (2007) Copula modeling: An Introduction for practitioners. *Foundations and Trends in Econometrics*, 1(1), Now Publishers, Inc., 2007.

U.S. Census Bureau (2009) Estimates of the annual rates of the components of resident population change for the United States, Regions, States, and Puerto Rico: July 1, 2008 to July 1, 2009, NST-EST2009-06, Population Division.

U.S. DOE (2008) Properties of fuel, U.S. Department of Energy. Available at: www.afdc.energy.gov/pdfs/fueltable.pdf

U.S. DOE (2012) Transportation Energy Data Book, Edition 31, U.S. Department of Energy. Available at: http://cta.ornl.gov/data/tedb31/Edition31_Full_Doc.pdf

Van Acker, V. and F. Witlox (2010a) Commuting trips within tours: how is commuting related to land use? *Transportation*, 38(3), 465-486.

Van Acker, V. and F. Witlox (2010b) Car ownership as a mediating variable in car travel behaviour research using a structural equation modeling approach to identify its dual relationship. *Journal of Transport Geography*, 18(1), 65-74.

Vance, C. and R. Hedel (2007) The impact of urban form on automobile travel: disentangling causation from correlation. *Transportation*, 34(5), 575-588.

Vega, A. and A. Reynolds-Feighan (2009) A methodological framework for the study of residential location and travel-to-work mode choice under central and suburban employment destination patterns. *Transportation Research Part A*, 43(4), 401-419.

Vovsha, P. and M. Bradley (2006) Advanced activity based models in context of planning decisions. *Transportation Research Record: Journal of the Transportation Research Board*, 1981, 34-41.

Waddell, P., C.R. Bhat, N. Eluru, L. Wang, and R. Pendyala (2007) Modeling interdependence in household residence and workplace choices. *Transportation Research Record: Journal of the Transportation Research Board*, 2003, 84-92.

Wen, C-H. and F.S. Koppelman (1999) Integrated model system of stop generation and tour formation for analysis of activity patterns. *Transportation Research Record: Journal of the Transportation Research Board*, 1676, 136-144.

Yamamoto, T. (2008) The impact of life course events on vehicle ownership dynamics-the cases of France and Japan. *IATSS Research*, 32(2), 34-43.

Yamamoto, T. and R. Kitamura (1999) An analysis of time allocation to in-home and out-of-home discretionary activities across working days and non-working days. *Transportation*, 26(2), 231-250.

Yamamoto, T. and R. Kitamura (2000) An analysis of household vehicle holding durations considering intended holding durations. *Transportation Research Part A*, 34(5), 339-351.

Yamamoto, T., R. Kitamura, and S. Kimura (1999) Competing risks duration model of household vehicle transactions with indicators of explanatory variables. *Transportation Research Record: Journal of the Transportation Research Board*, 1676, 116-123.

Ye, X., R.M. Pendyala, and G. Gottardi (2007) An exploration of the relationship between mode choice and complexity of trip chaining patterns. *Transportation Research Part B*, 41(1), 96-113.

Zhang, J., H.J.P. Timmermans, and A. Borgers (2005) A model of household task allocation and time use. *Transportation Research Part B*, 39(1), 81-95.

Zhao, Y. and H. Joe (2005) Composite likelihood estimation in multivariate data analysis. *The Canadian Journal of Statistics*, 33(3), 335-356.

Vita

Rajesh Paleti was born to his parents, Satya Seeta and Ashok Babu, in Eluru, India. He did most of his schooling in three cities- Hyderabad, Chennai, and Delhi. After completing high school at New Generation Junior College, Hyderabad, he joined Indian Institute of Technology, Kharagpur in the year 2004 to pursue undergraduate studies in Civil Engineering. He successfully graduated with a Bachelors degree in Civil Engineering in the year 2008. In Fall 2008, he joined University of Texas at Austin to pursue graduate studies in Transportation planning under the guidance of Dr. Chandra R. Bhat. He received his M.S. degree in Civil Engineering from The University of Texas at Austin in December 2009 and later continued for his doctoral degree under the guidance of Dr. Chandra R. Bhat.

Permanent address: 302 Ganapathi Homes
Phanigiri Colony, New Maruthi Nagar,
Chaitanyapuri, 500060, Hyderabad Andhra Pradesh, India

This dissertation was typed by the author.